Universidad Autónoma de Madrid

Universidad Autónoma de Madrid

Departamento de Biología Molecular

Facultad de Ciencias

Structural characterization of *Saccharomyces cerevisiae* mitochondrial transcription machinery

- TESIS DOCTORAL -

Srdja Drakulic

Madrid, 2013

CENTRO NACIONAL DE BIOTECNOLOGIA

"la Caixa"

CSIC

Universidad Autónoma de Madrid

Departamento de Biología Molecular

Facultad de Ciencias

Memoria presentada para optar al grado de Doctor en Ciencias Biológicas por

Srdja Drakulic

Universidad Autónoma de Madrid
Enero de 2013

DIRECTORES DE TESIS:
Dr. José María Valpuesta Moralejo
C.N.B.-C.S.I.C.
Dr. Jorge Cuellar Pérez
C.N.B.-C.S.I.C.

El trabajo recogido en esta memoria ha sido realizado en el Centro Nacional de Biotecnología (C.N.B.-C.S.I.C.) bajo la dirección conjunta de los Drs. José Maria Valpuesta Moralejo y Jorge Cuellar Pérez. Su financiación corrió a cargo de una beca del la Fundación "la Caixa" y programa para desarrollo doctorados en Biomedicina en el Centro Nacional de Biotecnología del CSIC

ACKNOWLEDGMENTS

These years have been filled with extraordinary scientific and human experience. I have met lots of amassing and warm people along the way. I would like to express my sincere gratitude to everyone who supported me and cared about me during my time in CNB.

Professor José María Valpuesta, for giving me the opportunity to work under your guidance and for a chance to work in science. Thanks for teaching me to stay focused to produce interesting results and to give my very best. I appreciate the faith and trust that you put in me and I have grown to respect you as a scientist and as a person. Thanks also for your endless patience regarding all the strange and unplanned things that occurred along the way.

Professor Rui Sousa, for involving us in the realization of this project and for fruitful and outstanding collaboration.

Jorge Cuellar, It's hard to choose where to start and how to thank you. I would like to thank you for accepting a total stranger as friend, without any prejudice. I am really lucky to find you. Thank you for not being tired of listening to all my problems and discouragements. Thank you for showing me what the real friendship is. I would like to show my gratitude for all the endless conversations that we had and for making me not to feel away from home. Without you this thesis wouldn't be possible.

Sensei, I don't know whether to be happy or to be worried for inevitably turning into you. Thank you for teaching me the processing and for having patience with me (and my non-existing Spanish).

Sara, for all the help and cloning lessons; **Josue, José, and Elena** for listening to all my crazy ideas and theories during the lunch brakes, and to all of you for friendship, warmth and encouragement over these years (especially you Josue).

Rocio, for showing me the interesting world of cryo-EM; **Cristina Patiño** and **Javi** for all the things that you showed me and an excellent job that you are doing in the SME.

All the people in the lab that I haven't mentioned, for an amazingly pleasant and friendly every working day. Thank you all for tolerating me and my creative mess.

I na kraju bih hteo da se zahvalim mojim roditeljima, sestri i baki na bezuslovnoj podršci i beskrajnoj ljubavi koji su mi pružili tokom svih ovih godina. Sve što sam postigao dugujem vama.

Mateji, mom malom i beskrajno strpljivom Sanchu Panzi na svim suludim stvarima koje smo proživeli u ovih 5 godina u tuđini. Hvala za svaki nezaboravni trenutak, konstantnoj potpori i neiscrpnoj ljubavi.

Live long and prosper!!!!

ABSTRACT

Mitochondria play important roles in a plethora of cellular processes such as energy production, cell growth, cell differentiation, programmed cell death (apoptosis), signaling and innate immunity. Therefore, they are an essential part of every living eukaryotic cell. Due to the fact that mitochondria are semi-autonomous organelles, genes involved in the number of the mitochondrial processes are localized in mitochondria and nucleus. Transcription, the first and the most critical step in gene expression, requires the presence of complex molecular machines, DNA-dependent RNA polymerases, which can be divided in two groups: (1) Multi-subunit RNAPs, found in eubacteria, archaea and in the nucleus of eukaryotic organisms; and (2) single-subunit RNAPs, localized in some bacteriophages, in the mitochondria and chloroplast of eukaryotic organisms, and in the nucleus of mammalian cells. The mitochondrial transcription has its roots in the interaction of three entities: the host cell, the bacteria and phage, thus, sharing a number of features with their transcription systems. Even though mtRNAPs are often placed under single-subunit RNAP family, their activity is dependent and regulated by a number of different factors and cofactors.

This thesis is focused on the structural characterization of the *Saccharomyces cerevisiae* mitochondrial transcription machinery, which consists of the core RNA polymerase (sc-mtRNAP) and an auxiliary transcription factor, MTf1. For the first time the three dimensional reconstructions of full length sc-mtRNAP in quasi-IC (bubble promoter:sc-mtRNAP complex), IC (bubble promoter:sc-mtRNAP complex:Mtf1) and EC (bubble DNA:RNA:sc-mtRNAP complex), together with the N-terminal deletion mutant EC, have been performed, and the DNAs' termini have been localized in those complexes. The structural data presented in this thesis also confirms the existence of the differential DNA bending that was previously shown to be responsible for discrimination of promoter from non-promoter sequences. Furthermore, our data indicate that sc-mtRNAP undergoes similar structural rearrangement as T7 RNAP, where the large conformational changes are constrained only to the N-terminal domain and the upstream region of DNA. Thus, the thesis has served for filling in the existing information gap regarding the structure of sc-mtRNAP, which would help in deciphering the mechanisms that are in the basics of the yeast mitochondrial transcription.

TABLE OF CONTENTS:

FIGURE INDEX:

ABRREVATIONS:

2D: two-dimensional

3D: three-dimensional

3DEM: three-dimensional electron microscopy

A: adenosine

Å: Angstrom

Abf2: abundant high-mobility group DNA binding protein

ATP: adenosine-5'-triphosphate

bp: base pair

C: cytosine

CAP: catabolite activator protein

CL2D: clustering two-dimensional classification

cryo-EM: cryo-electron microscopy

CTD: carboxyl terminal domain

CTF: contrast transfer function

Cu/Rh: Copper/Rhodium

DBS: double-stranded deoxyribonucleic acid -binding site

DMT: deoxyribonucleic acid methyltransferase

DNA: deoxyribonucleic acid

DNAPI: deoxyribonucleic acid polymerase I

DTT: dithiothreitol

EC: elongation complex

EDTA: ethylenediaminetetraacetic acid

eV: electron volts

F: force

Fe-BABE nuclease: Fe (III) (S)-1-(p-Bromoacetamido-benzyl)ethylene diamine tetra acetic acid nucleases

FSC: Fourier Shell Correlation

G: guanine

GTP: guanosine-5'-triphosphate

H_2O_2: hydrogen peroxide

HBS: hybrid binding site

HCl: Hydrochloric acid

HhH: helix–hairpin-helix motifs

HIV I RT: human immunodeficiency virus I reverse transcriptase

h-mtTFB1: Human mitochondrial transcription factor

Ilv5: alpha-keto-beta-hydroxylacil reductoisomerase

IPTG: Isopropyl-1-thio-β-D-galactopyranoside

rut: utilization site

kDa: kilo Dalton

KerDenSom: Kernel Density Estimator Self-Organizing Map

LB medium: lysogeny broth medium

M: molar

Mdm10p: mitochondrial distribution and morphology protein 10

Mdm12: mitochondrial distribution and morphology proteins 12

ML2D: maximum likelihood two-dimensional classification

ML3D: maximum likelihood three-dimensional classification

mRNA: messenger : ribonucleic acid

mtDNA: mitochondrial deoxyribonucleic acid

mtRNAP: mitochondrial ribonucleic acid polymerase

MWM: molecular weight marker

NaCl: sodium chloride

NTD: amino terminal domain

NTP: nucleoside triphosphate

nts: nucleotides

PBD: promoter binding domain

PEI: polyethyleneimine

PMSF: phenylmethanesulfonylfluoride

PPi: pyrophosphate

RBS: ribonucleic acid binding strand

RNA: ribonucleic acid

RNAP I: ribonucleic acid polymerase I

RNAP II: ribonucleic acid polymerase II

RNAP III: ribonucleic acid polymerase III

RNAP: ribonucleic acid polymerase

Rpb 1-11: ribonucleic acid polymerase II subunit 1-11

RPc: initial closed promoter complex

RPitc: scrunched initiation complex

rpm: rotations per minute

RPo: open promoter complex

rRNAs: ribosomal : ribonucleic acid

RT: room temperature

sc-mtfB: *Saccharomyces cerevisiae* mitochondrial transcription factor B

sc-mtRNAP: *Saccharomyces cerevisiae* mitochondrial ribonucleic acid polymerase

SDS-PAGE: sodium dodecyl sulfate polyacrylamide gel electrophoresis

snRNA: small nuclear ribonucleic acid

T: thymidine

T7 RNAP: T7 phage ribonucleic acid polymerase

TRIS: tris(hydroxymethyl)aminomethane

tRNA: transport : ribonucleic acid

tsp: transcription stop points

T-stretch: oligo thymidine sequence

TTF1: transcription termination factor 1

UP element: promoter upstream regulatory element

wt: wild type

µg: micro grams

µl: micro liters

Z: atomic number

AMINO ACIDS, ONE AND THREE LETTER CODES

Amino acid	Three letter code	One letter code
alanine	ala	A
arginine	arg	R
asparagine	asn	N
aspartic acid	asp	D
cysteine	cys	C
glutamic acid	glu	E
glutamine	gln	Q
glycine	gly	G
histidine	his	H
isoleucine	ile	I
leucine	leu	L
lysine	lys	K
methionine	met	M
phenylalanine	phe	F
proline	pro	P
serine	ser	S
threonine	thr	T
tryptophan	trp	W
tyrosine	tyr	Y
valine	val	V

1. Introduction

1.1. Flow of genetic information

Almost 80 years after Mendel's discovery of hereditary nature of the traits in garden peas (Mendel, 1886), experiments by Avery and coworkers settled down a major dispute in scientific community regarding the chemical nature of genetic material (Avery *et al.*, 1944). Until that moment, the large majority of scientists strongly believed that proteins, rather than nucleic acids, represent a material form of genetic information, mainly due to their higher complexity. These experiments, together with the ones by Hershey and Chase, clearly showed that the nature of genetic material is always the nucleic acid (Hershey y Chase, 1952), imprimis deoxyribonucleic acid (DNA), with the exception of ribonucleic acid viruses (RNA viruses). The transfer of genetic material through the generations is based upon double helical structure of DNA. Through the process of DNA duplication, known as DNA replication, daughter strands are synthesized upon the mother strands. Consequently, every progeny cell contains a complete DNA molecule or a set of DNA molecules in organisms whose genomes contain more than one chromosome. Even though the replication in principle is straightforward, the process is extremely complex and requires the presence of a multitude of cellular factors, in order to be performed with fidelity and efficiency (Lewin, 2003).

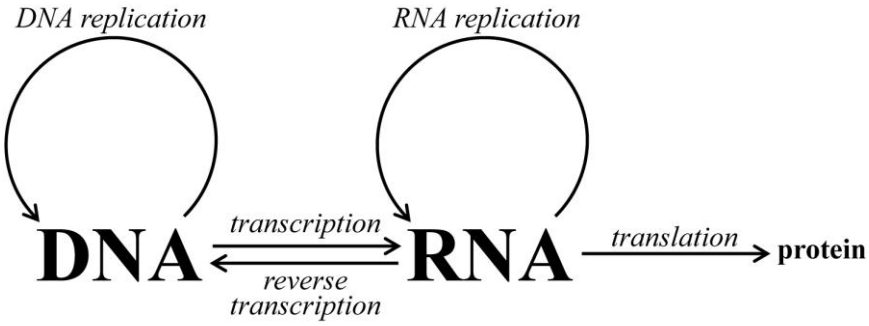


Figure 1: Schematic illustration of Central dogma of molecular biology adapted from (Crick, 1970). One-dimensional information, stored as a sequence of nucleotides in DNA, is converted to three-dimensional (3D) structure, protein. The conversion is not a direct, one-step process, rather involves an intermediate product, RNA. Thus, two steps can be distinguished: transcription and translation. Scheme also illustrates the existence of a genetic information transfer from RNA to RNA (RNA replication), and from RNA to DNA (reverse transcription) present in RNA viruses.

Despite the fact that amino acid sequence, in proteins, is defined by the order of triplets of bases in a DNA molecule, DNA is not the direct template for protein synthesis. This gap in information flow from DNA to protein is bridged by RNA molecules, which are created through the process of transcription (Figure 1). Protein encoding genes are transcribed into the messenger RNAs (mRNA), which in the case of eukaryotes are immediately subjected to the process of maturation and subsequent transport from the nucleus to the cytoplasm. Protein synthesis, named translation, occurs in the cytoplasm through a complementary pairing of nucleotide triplets between messenger and transport RNAs (tRNA). The whole process is catalyzed by the ribosome, a large and complex ribonucleoprotein machine, made of two main subunits. The subunit composition differs between prokaryotes and eukaryotes, but generally includes one to several ribosomal RNAs (rRNAs) and a large number of small ribosomal proteins (McQuillen *et al.*, 1959).

1.2. Transcription

Transcription represents the first step in gene expression, in which a RNA molecule, identical to the coding strand (also called non-template, plus or sense strand), and complementary to the template strand (also known as minus or antisense strand), is synthesized. The growth of RNA is always in the 5'→3' direction, while the template strand is read in the opposite, 3'→5' direction. The first nucleotide at the 5' end retains the triphosphate group. Process of RNA synthesis is performed by RNA polymerase (RNAP, see the next heading), which starts the transcription by binding to a specific region, the promoter, at the beginning of a gene. A promoter is a regulatory DNA sequence, which determines when, where and to what extent a gene will be expressed. Its composition drastically varies from prokaryotes, through archaea to eukaryotes. In general, promoters contain a startpoint, the first base that is transcribed into RNA, and regulatory elements, recognized by RNAP either alone or in complex with transcription factors. From the startpoint, the RNAP moves along the template strand until reaches the terminator sequence. The sequence between the promoter and the terminator is defined as transcription unit, which may include more than one gene. The first product of transcription is called the primary transcript, and in most of the cases is highly unstable. In prokaryotes it is rapidly processed to generate the mRNA or cleaved to render mature products in the rRNAs and the tRNAs. In eukaryotes, the primary transcript undergoes a maturation process, through modifications at its ends (in the case of the mRNA) and/or cleaving (in all the RNAs).

Generally, the entire transcription process can be divided into 3 stages:

- **Initiation:** includes recognition, binding and promoter melting, as well as synthesis of short RNAs, forwhich this step is also known as abortive synthesis. RNAP remains bound to the promoter during these initial steps, while the growing RNA chain probably occupies its active site. If the polymerase manages to release the promoter, in a same time without losing the growing RNA chain, and to move the next region of the template to the active site, the initiation is accomplished. Otherwise, the RNAP loses the contacts with the RNA which aborts the initiation, and the whole process has to be repeated.
- **Elongation:** The RNAP forms a stable elongation complex (EC), in which downstream DNA duplex (incoming DNA) is unwound to form a transcription bubble. The transcription bubble contains the DNA-RNA hybrid. Behind the unwound region (upstream DNA), two DNA strands pair to reform the double helix. In general, RNAPs in elongation establish more extensive contacts with DNA-RNA hybrid, contributing to a higher stability of the EC.
- **Termination:** involves recognition of a terminator, the point at which new bases should not be added to the RNA. This is followed by a collapse of the transcription bubble, and the release of the enzyme and the RNA. Termination employs different mechanisms. One of them is to acquire the specific secondary structure in synthesized RNA consisting of palindrome sequence, rich in guanine (G) and cytosine (C) that forms a stem-loop hairpin in bacterial primary transcript. Other strategy is the employment of termination factors, such as Rho terminator in bacteria that binds to

5′ end of RNA or transcription termination factor 1 (TTF1) in the case of eukaryotic RNAPI, which interacts with specific sequence downstream of the transcription unit. Finally, in the case of bacterial RNAP and eukaryotic RNAPIII, specific arrays of nucleotides are used for termination: thymidine (T) and thymidine-adenosine (T-A) stretches, respectively.

1.3. RNA polymerases

In all organisms gene expression requires the presence of complex molecular machines, DNA-dependent RNA polymerases, which can be divided in two groups: (1) Multi-subunit RNAPs, found in eubacteria, archaea and in the nucleus of eukaryotic organisms; and (2) single-subunit RNAPs, localized in some bacteriophages, in the mitochondria and chloroplast of eukaryotic organisms, and in the nucleus of mammalian cells. These two groups represent different mechanisms or approaches to tackle the problem of RNA synthesis, which is that the dissociation of the polymerase from a template has to be much slower than the adding of new bases to the growing, newly synthesized RNA. Moreover, the polymerase has to be able to translocate along the template. From a structural point of view, the nature has reached the solution by devising the polymerases capable of templates' encircling. In the case of the multi-subunit RNAPs, the problem is overcome by the formation of a doughnut-like structure. The single-subunit RNAPs, however, have evolved flexible structures, capable of wrapping around the template. The structural studies of the Klenow fragment, a large proteolitic product of DNA polymerase I (DNAPI), with 5'-3' polymerase and 3'-5' exonuclease activities, of the human immunodeficiency virus I reverse transcriptase (HIV I RT) and of the T7 phage RNAP (T7 RNAP) have shown the existence of this flexible structure named thumb domain, which stabilizes the complex with the template during the processive transcription (Bonner *et al.*, 1994, Sousa *et al.*, 1994).

1.3.1. Multi-subunit RNA polymerases

In archaea and bacteria, the synthesis of all transcriptional RNAs is accomplished by a unique multi-subunit RNAP. However, eukaryotes have diversified this task, employing three different multi-subunit RNAPs. Thus, RNAP I produces rRNAs, while RNAP II and III are involved in synthesis of mRNAs and tRNAs, respectively. RNAP II is also responsible for the production of small nuclear RNAs (snRNA). The sequence analysis showed the evolutionary conservation of these multi-subunit RNAPs, with the total of 22 homology regions located in the core subunits cluster around the active site (Cramer, 2004).

In the past couple of years, the structures of archaeal (*Sulfolobus salfataricus* and *Pyrococcus furiosus*), bacterial (*Thermus aquatics, Thermus thermophilus*) and eukaryotic nuclear RNAPs (*Saccharomices cerevisiae*) have been determined, revealing a common architecture of the core whose shape is similar to a crab's claw (Zhang *et al.*, 1999, Vassylyev *et al.*, 2002, Cramer, 2004, Vassylyev *et al.*, 2007b, Cramer *et al.*, 2008, Hirata *et al.*, 2008, Kusser *et al.*, 2008, Hirata y Murakami, 2009, Cheung *et al.*, 2011, Martinez-Rucobo y Cramer, 2012, Treutlein *et al.*, 2012, Wu *et al.*, 2012). The structural comparison of

these multi-subunit RNAPs has shown the presence of many structurally homologous regions, not only the ones pointed out by sequence analysis, which suggest that a general structure has been conserved during evolution (Cramer, 2004).

The central mass of the enzyme and opposite sides of the positively charged cleft are formed by two large subunits in bacteria and eukaryotes, β and β' in bacterial RNAP; RNA polymerase II subunit 1 and 2 (Rpb1 and Rpb2) in eukaryotic RNAP (Cramer, 2002). In archaea, this large ellipsoidal body is composed of three subunits A', A'' and B, where A' and A'' are homologous to the amino (N-) and carboxyl (C-) terminal parts of β' and Rpb1 in bacteria and yeast RNAP II, respectively (Figure 2). The central mass is formed and held together on the assembly platform, composed of two small core subunits (α homodimer, Rpb3-Rpb11 heterodimer, and D-L heterodimer in bacteria, eukaryotes and archaea, respectively) (Figure 2). In archaeal RNAP and eukaryotic RNAP II, the assembly platforms are extended to two additional polypeptides (N and P, and Rpb10 and Rpb12, respectively) (Werner *et al.*, 2000).

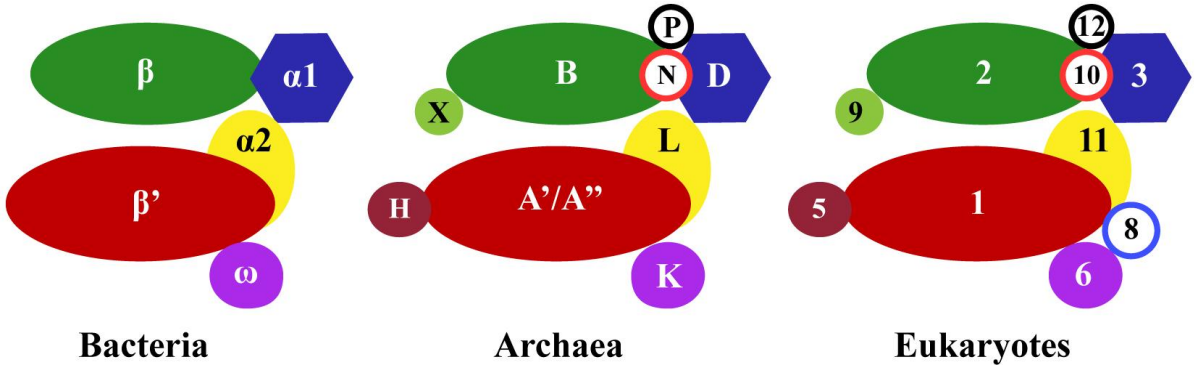


Figure 2: Schematic presentation of multi-subunit RNAP structure in three domains of life: Bacteria, Archaea and Eukaryotes. Subunits that exhibit sequence and structural homology have the same color. Overall structure resembles a crab´s claw. The two pincers are represented in red (β', A'/A'' and Rpb1 (1)) and green (β, B, and Rpb2 (2)); The structure is held together by the assembly platform, which comprises α1 and α2 in Bacteria; P, N, D and L in Archaea and Rpb12 (12), Rpb10 (10), Rpb3 (3), Rpb11 (11) and Rpb8 (8) in Eukaryotes. Adapted from (Cramer, 2002)

The lower part of the claw (β', A'/A'', Rpb1 side) consists of a flexible, mobile element, known as clamp. It contains an extra unit in archaea and eukaryotes (H and Rpb5, respectively), which are structurally and functionally mimicked by an insertion of β' subunit in bacteria (or *vice versa*). These elements have been shown to interact with the downstream DNA (Bartlett *et al.*, 2004). The active site is located at the base of the cleft, close to the enzyme's center (Figure 2). The other side is divided in 2 domains: the lobe and the protrusion in RNAP II, and β domains 1 and 2 in bacterial RNAP. The cleft is closed on one side by the wall or the flap. The structurally conserved core includes the active site, which indicates that all multi-subunit RNAPs share a common mechanism (Cramer, 2002).

1.3.1.1. Bacterial RNAP holoenzyme structural organization

Bacterial RNAP is a holoenzyme with molecular mass of around 465 kDa, consisting of catalytically active core (~400kDa) and an extra protein called sigma (σ) factor, whose presence is required for the accurate promoter recognition, binding and melting. The core has a crab´s claw shape, with the main mass

formed by the β and the β' subunits, as mentioned before. The smaller, top pincer is made of β, while the β' comprises the larger, bottom pincer (Figure 3A). The inner surface of the cleft is rich in positively charged amino acids, while the outer surface of RNAP is mainly negatively charged (Figure 3B).

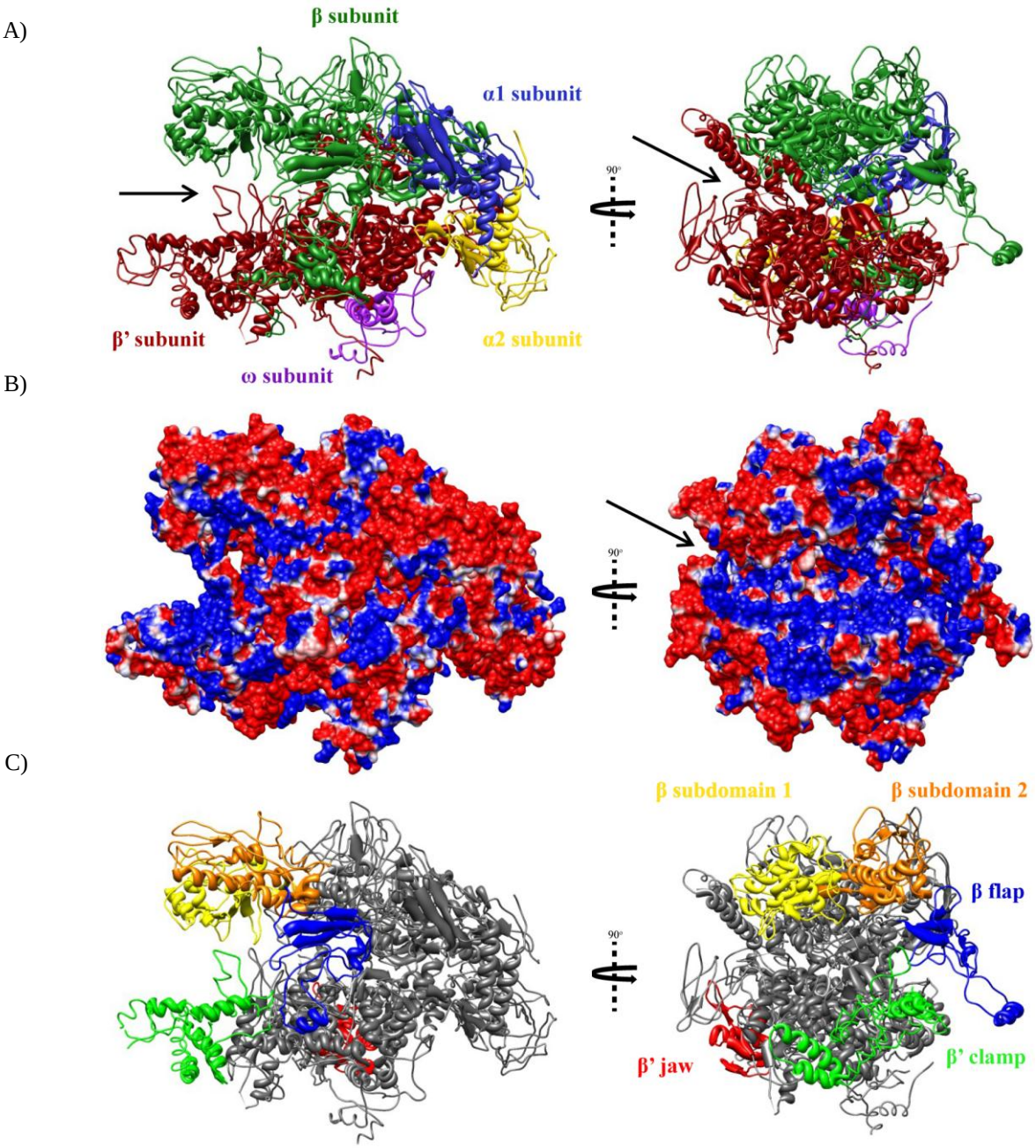


Figure 3: Bacterial RNAP structure. A) Bacterial RNAP subunit composition. X-ray structure of *T. aquaticus* RNAP (Protein Data Bank (PDB) reference 1I6V), resembling crab´s claw, with β (green), β' (red), α1 (blue), α2 (yellow), and ω (purple) subunits, is shown in two standard views (side and top). The binding cleft is indicated by the black arrows. Color code is the same as for the figure 2. **B)** Distribution of surface electrostatic charges. The positive charges (blue) are predominantly localized in the binding cleft (indicated by the black arrow), involved in interaction with the nucleic acids, while the negatively charged amino acids (red) mostly comprise the outer surface. **C)** Mobile elements of core RNAP: β lobes (subdomain 1 and 2; yellow and orange), β flap (blue), β' jaw (red), β' clamp (green).

Despite the fact that α subunit is present in a dimeric form, these two subunits have non-equivalent functions (Zhang *et al.*, 1999). Therefore, α1 and α2 subunits can be distinguished. α1 is only in contact with the β, while α2 exclusively makes contacts with β' (Figure 3A). The α dimer functions as the assembly platform, which is formed first, and subsequently followed by the binding of the β and β' subunits (Ishihama

et al., 1987). Further, each of α subunits can be divided in 2 functional domains bridged by a flexible linker: (1) N-terminal domain (NTD), which plays a role in enzyme assembly and is located on the surface of the enzyme, opposite from the deep cleft; (2) C-terminal domain (CTD), which is involved in binding to the promoter upstream regulatory element (UP element) and interaction with transcriptional activators. Structural studies have shown that αCTD is a compactly folded domain with four α helices and one non-standard helix, forming two helix–hairpin-helix (HhH) motifs (Jeon *et al.*, 1995, Bar-Nahum y Nudler, 2001). CTD achieves multiple functional interactions with the DNA in the UP element, with activators such as catabolite activator protein (CAP), and with the region 4 of σ factor. This suggests that the interaction between the CTD and the σ is involved in transcription activation, possibly, through the stabilization of σ region 4's contact with the -35 promoter element (Geszvain y Landick, 2012).

The multi-subunit core of bacterial RNAP is perfectly capable of performing the elongation and termination steps of transcription, but the promoter-specific initiation requires σ factor that directs RNA polymerase to the specific promoter elements and subsequently plays a key role in double stranded DNA (ds DNA) melting, and therefore transcription bubble formation. Due to this crucial role, σ factors represent the key point in gene expression control. The availability and activity of σ factors are regulated by, at least, two types of regulatory factors: anti- σ factors, which bind and inhibit σ factors, and appropriators that act on specific RNAP holoenzyme, altering their activity (Campbell *et al.*, 2005, Campbell *et al.*, 2008). All known σ factors fall in two broad classes (Geszvain y Landick, 2012):

- σ^{70} family or σ1 group, where the majority of known sigma factors belong. The members of this group are similar to *Escherichia coli*'s housekeeping σ or σ^{70} factor and they are present in all bacterial species. Based on their gene structure and the involvement at different stages of bacterial growth cycle, the members of σ^{70} family members can be further subdivided into four subgroups (Schwartz *et al.*, 2008).
- σ2 group, whose members share a similar structure with σ^{54} or σ^{N}, responsible for transcription of the genes implicated in nitrogen fixation and stress response.

Sequence analysis and structural studies of the σ^{70} family has revealed the existence of four conserved regions (σ1.1, σ2, σ3, σ4), of which only 2 and 4 are well preserved and present in all members of this family (Malhotra *et al.*, 1996, Campbell *et al.*, 2002, Schwartz *et al.*, 2008). The σ2, σ3, σ4 consist of several conserved subregions (Figure 4A) which are compactly folded domains joined by flexible linkers (Figure 4B) (Campbell *et al.*, 2002, Schwartz *et al.*, 2008). The presence of flexible linkers allows the correct positioning of subregions during the initiation of transcription.

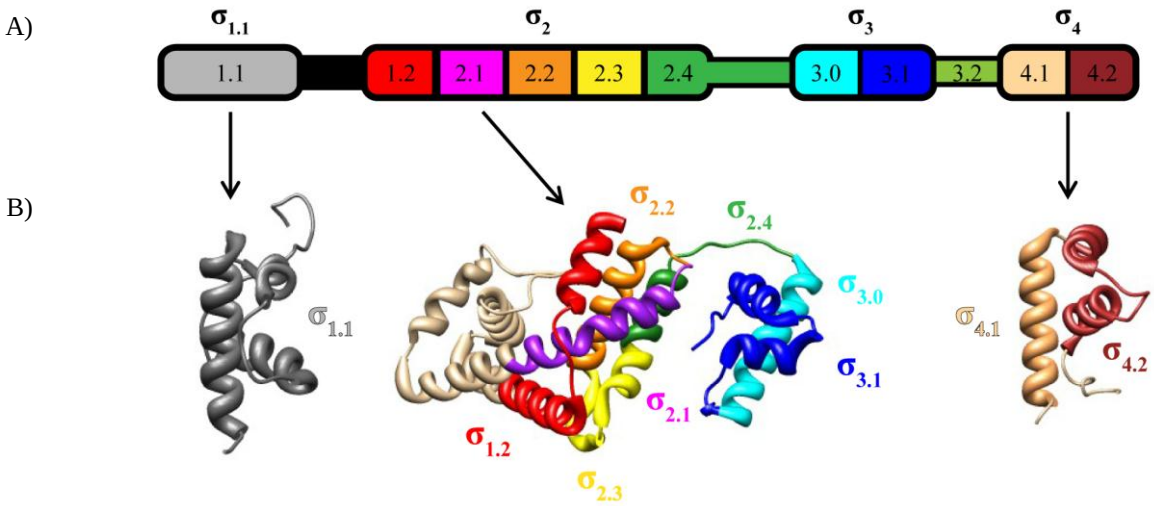


Figure 4: Conserved regions of σ^{70} family and the structure of corresponding proteolytic fragments. **A)** Schematic representation of evolutionary conserved regions and comprising subregiones. The conserved subregiones are color coded: σ1.1 (grey); σ1.2 (red); σ2.1 (purple); σ2.2 (orange); σ2.3 (yellow); σ2.4 (green); σ3.0 (cyan); σ3.1 (blue); σ4.1 (beige); σ4.2 (brown). **B)** Structures of proteoliticly obtained fragments of σ1.1 (*Thermus maritima*, PDB 2K6X), regions σ2-3 (*Thermus aquaticus* σ^A, PDB 1KU2) and region σ4 (*T. aquaticus* σ^A, PDB 1KU3). The color code was maintained from the figure b).

σ1.1 domain is a three helical domain characterized by a high acidic content that makes this region mostly negatively charged. This is important for establishing the electrostatic interactions with positively charged σ2 and σ4 regions in σ free factor. As a consequence, a free σ factor has a compact structure, where the σ1 locks the DNA binding domains and prevents their interaction with the DNA promoter (Dombroski *et al.*, 1992, Dombroski *et al.*, 1993, Camarero *et al.*, 2002). However, as it has been previously shown, these interactions must not be excessively strong, otherwise it would be energetically expensive to form the RNAP holoenzyme capable of binding to the DNA (Dombroski, Walter et al. 1993). Through the interaction with the catalytic core, it is believed that the σ1.1 is separated from the rest of the σ factor. The X ray structures of σ2 domain revealed a bundle of three α-helices that play important role in holoenzyme formation through the interaction with coiled-coil domain of β' pincer (Malhotra *et al.*, 1996, Campbell *et al.*, 2002, Li *et al.*, 2002). It has been also shown that conserved regions σ2.3 and σ2.4 are implicated in recognition of -10 element (Murakami *et al.*, 2002a) and DNA promoter melting (Murakami y Darst, 2003). The σ3 domain is less conserved among the members of the the σ70 family. This is a three-helix compact domain, in which residues of the first N-terminal helix (σ3.0) are implicated in the recognition of the extended -10 element (Paget y Helmann, 2003). The σ4 is a C-shaped domain with two distinguishable subregions, σ4.1 and σ4.2, responsible for the core RNAP binding and for making the initial contacts with the -35 element of the promoter (Campbell *et al.*, 2002, Murakami y Darst, 2003).

1.3.1.2. Bacterial RNAP transcription cycle

As mentioned before, transcription is a site-specific cyclic process that can be divided into three major steps, each of them consisting of multiple smaller ones. The initiation is the most complex and the most critical one. In bacteria, the first initiation step is the formation of the functional holoenzyme (Figure 5). Namely, the second helix of the σ2 domain makes the principal interaction with coiled-coil domain of β'

pincer (β' major docking site), orientating the σ3 domain and the remaining part of the σ2 domain towards the β subdomain 1 and the β subdomain 2, respectively. Additionally, the hydrophobic residues of σ4.1 subregion form the surface of pocket in which the β flap-tip helix of core is inserted, building the σ4- flap mobile module and positioning the σ4 near the exterior of the RNA exit channel (Vassylyev *et al.*, 2002). The N-terminal part of the σ3-σ4 linker forms a hairpin loop, which is buried inside the RNAP major channel. Its C-terminal part, or the σ3.2 region, lies within the RNA exit channel (Mekler *et al.*, 2002). It is believed that, simultaneously, σ1.1 gets separated from the rest of the σ factor, which releases the factor (in a first place σ2 and σ4 domains) of its inhibitory effect (Schwartz *et al.*, 2008).

Once the holoenzyme is formed and released from the σ1.1's inhibitor effect, it can interact with the DNA, forming the initial closed promoter complex (RPc) (Figure 5). The σ2 and σ4 domains make contacts with -10 (^{12}TATAAT^{7}) and -35 (^{35}TTGACA^{30}) elements, respectively. The conserved σ4.2 region establishes the contacts with the region spreading from -38 to -30, which induces the DNA bending at the -35 element. This brings the upstream DNA closer to RNAP and allows the interaction between the αCTD and the UP element, contributing to the stabilization of the RPc (Bose *et al.*, 2008). At the same time, conserved residues of σ2.4 are placed within the reach of -12 position. All these events allow the holoenzyme and its elements to acquire an adequate position and a correct orientation towards the DNA. In the RPc, the DNA remains in a double stranded state and only partially protected from -50 to +5 position (Borukhov y Nudler, 2008). The transition to the open complex (RPo***)*** involves the interaction of perfectly positioned aromatic residues of σ2.3 with transiently exposed non-template bases of -10 element (Murakami *et al.*, 2002a). The natural breathing of the AT-rich element is in the origin of the single strand exposure, which could be stabilized by interactions with σ2.3. This leads to the formation of an initial, short segment of transcription bubble (Mekler *et al.*, 2002) (Murakami *et al.*, 2002a). Further unwinding from -11 to -7 would introduce flexibility in the DNA and lead to the entering of approximately 20 bps long downstream DNA duplex into the DNA-binding clamp. At this point, the transcription bubble consists of unwound DNA duplex from -11 to +2 (Borukhov y Nudler, 2008). Taken together, this results in correct placement of +1 base of the template strand in the catalytic center, allowing the start of initiation phase of transcription (Figure 5). The enzyme-catalyzed reaction begins by the placement of an adequate nucleotide triphosphate (NTP) at +2 position and formation of the first phosphodiester bond with the substrate at +1. It is followed by the release of pyrophosphate (PPi) and displacement of the nascent, growing RNA together with the DNA by one nucleotide towards the RNA exit channel. This sequence of actions is repeated until the RNA reaches the length of ~ 7- 10 nucleotides (nts) (Borukhov y Nudler, 2008). At the same time, the promoter remains bound to the σ factor with -10 and -35 elements, resulting in scrunching of the DNA (scrunched initiation complex (RPitc)) (Chen *et al.*, 2010). During these initial steps the RNAP holoenzyme faces two obstacles, which successful tackling leads to the formation of a highly processive elongation complex (EC) (Figure 5). Firstly, the scrunching of the DNA inevitably results in accumulation of free energy that eventually overcomes the one of the promoter binding. This has a consequence either the rupture of the contacts between the RNAP and the σ factor, and transition to elongation or breakage of the DNA-RNA hybrid, and return to the beginning. The latter occurs more

frequently. Secondly, as the nascent RNA reaches the length of ~ 11- 15 nts, its 5´ end encounters and collides with the σ3-4 linker and the σ4, which occupy the main and the RNA exit channel, respectively (Murakami *et al.*, 2002b). The clashing results either in disassembly of the DNA-RNA hybrid or in displacement of corresponding σ domains from the RNAP, in the first place the σ3.2 loop (Cashel *et al.*, 2003). This destabilizes the interactions between σ4 and the β flap, allowing the release of the -35 promoter element and the subsequent promoter escape (Kulbachinskiy y Mustaev, 2006). The σ factor, however, can remain bound to the RNAP, as the interactions between the σ2 and the RNAP are not affected (Borukhov y Nudler, 2003).

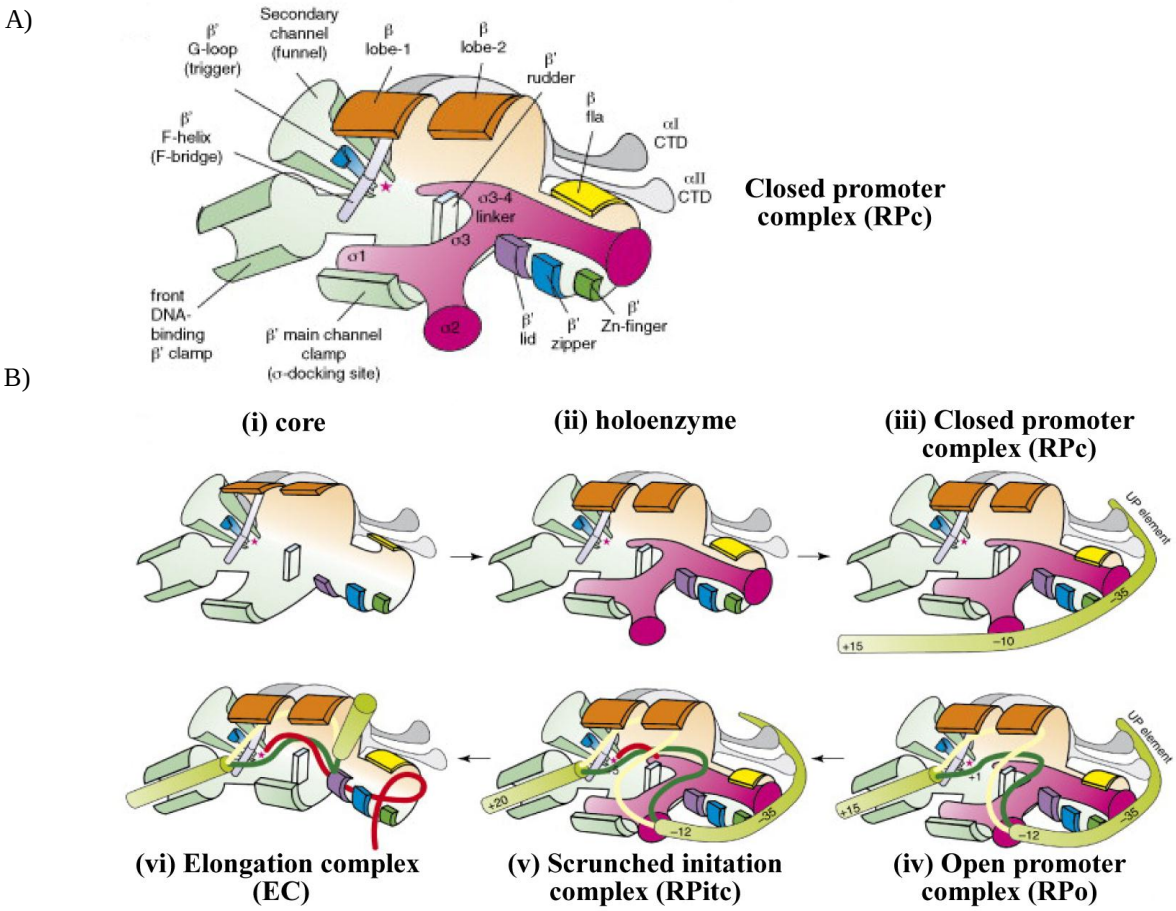


Figure 5: Illustration of bacterial RNA polymerase holoenzyme (*T. thermophilus*) and schematic diagram of the major RNAP transition steps during the transcription cycle from core to the elongation complex. A). The *T. thermophilus* RNAP holoenzyme is represented schematically. Mobile elements are color coded: β' F-bridge (purple); β' G-loop (blue); β' zipper (dark blue); β' Zn-finger (green); β' lid (light blue); β' rudder (turquoise); β lobe 1 and 2 (orange); and b flap (yellow). β' lid and β' rudder comprise mobile element named β' clamp in text. **B)** Keeping the same schematic representation of RNAP holoenzyme, the major RNAP transition steps during the transcription cycle from core to the elongation complex are displayed. Color code is the same as in A). Duplex DNA is shown as flexible green cylinders; nascent RNA, template and non-template DNA strands are shown as red, dark green and yellow strands, respectively. Movements of several mobile elements are exaggerated to emphasize probable conformational changes in RNAP: (ii) during σ binding to core (β lobes 1 and 2, β flap and β' lid can all move to accommodate and seal σ in holoenzyme); (iii) during promoter recognition to form RPc (β flap moves together with σ4 for its optimal placement towards -35 promoter region, and aCTD-I and –II move to bind UP element); (iv) during promoter melting in RPo (DNA binding clamp presumably closes on downstream duplex DNA to help formation of salt-resistant binary complex); (v) during abortive initiation in RPitc (DNA strands in the upstream part of transcription bubble are thought to be looped out to accommodate downstream DNA into the main channel during scrunching); and (vi) during promoter escape and transition to EC, β flap, β' zipper and β' lid first move to open the RNA exit channel to allow release of σ3, σ3–σ4 linker and σ4 from RNAP; and then together with β' rudder and β' main channel clamp they should move back to close the exit channel, stabilize the RNA–DNA hybrid and secure the placement of nascent RNA into the exit channel. Adapted from (Borukhov y Nudler, 2008)

The elongation step is characterized by a high stability, processivity and fidelity. The higher stability is obtained through the tighter main channel, increasing the surface for interaction with the nucleic acids (Vassylyev *et al.*, 2007a). Nevertheless, there are very few direct polar interactions between the protein and the downstream DNA, which remains highly mobile. This is also observed for the T7 RNAP EC, implying the downstream DNA mobility as the general feature responsible for the high processivity (Borukhov y Nudler, 2008). In the active center, bacterial RNAP contains two mobile parts essential for catalysis and RNA elongation: F bridge (bridge helix) and G loop (trigger loop). G loop exhibits high flexibility and can adopt many different conformations. In the closed or folded state, together with the F bridge, it is involved in selecting and binding of the adequate, correct NTP at i+1 site. This is achieved through the numerous established contacts between the G loop and the NTP(s). Upon the phosphodiester bond formation between the i-1 and the i+1 NTPs, those interactions are lost and the G-loop adopts an unfolded conformation. It has been indicated that the G-loop plays a crucial role in transcription fidelity (Borukhov y Nudler, 2008). Elongation is characterized by a continuous oscillation of EC between the active (elongation-competent), with the i+1 site ready to bind a substrate (NTP), and the inactive (backtracked) conformation, with the i+1 site occupied by the RNA (Borukhov y Nudler, 2008) (Figure 6).

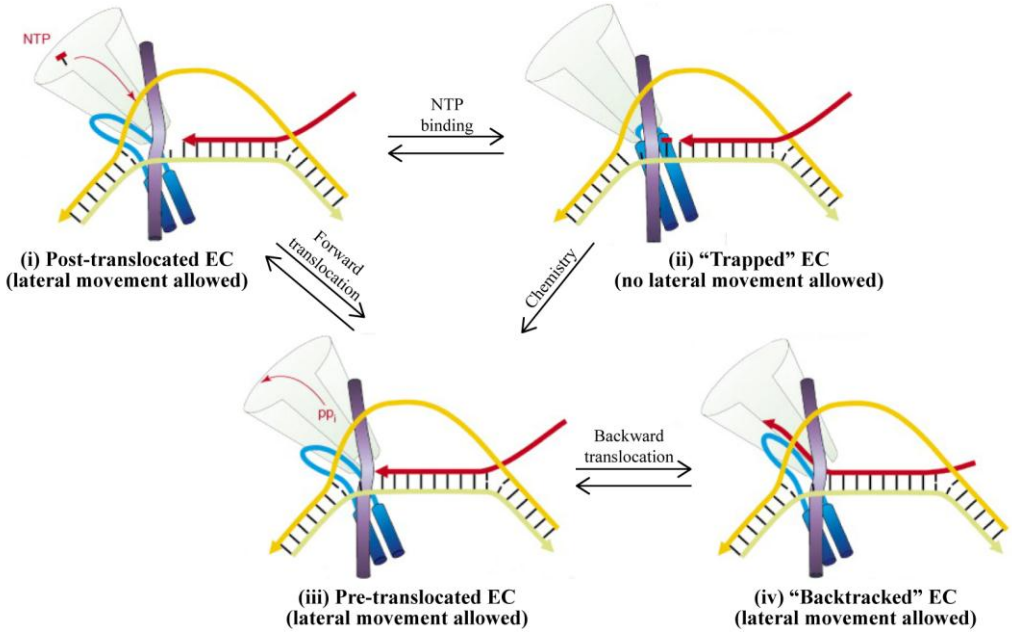


Figure 6: Dynamics of the catalytic center. Schematic illustration of the probable conformational changes in G(trigger)-loop and F(bridge)-helix elements of RNAP during NTP incorporation and subsequent translocation. The RNAP secondary channel is shown as a light cyan funnel; DNA template, non-template, and RNA strands are shown as thick green, orange and red strands, respectively, with arrowheads at the 3′ termini. DNA–DNA and DNA–RNA base pairing is indicated by small black bars. The β' F(bridge)-helix is shown as a purple column and the β' G(trigger)-loop is represented as two blue cylinders connected by a mobile loop. The diffusion of substrate NTP and pyrophosphate (PPi) through the secondary channel is indicated by the thin red curved arrow. In the initial active EC (i), G(trigger)-loop (together with the F(bridge)-helix) is in unfolded 'relaxed' conformation, which presumably allows lateral movement of RNAP. Binding of the correct NTP at the i+1 site restricts the lateral movement of RNAP and leads to a refolding of a 'relaxed' G(trigger)-loop–F(bridge)-helix unit conformation into an extended α-helical conformation (ii), which in turn further stabilizes NTP binding through interactions with the G-loop, and results in the formation of a 'trapped' EC. In the 'trapped' conformation, the G(trigger)-loop not only prevents the movement of RNAP but also obstructs the secondary channel passage. After incorporation of NTP into RNA and the release of pyrophosphate, the G(trigger)-loop–F(bridge)-helix unit reverts to 'relaxed' conformation once again, allowing the lateral movement of RNAP in pre-translocated EC (iii), resulting either in RNAP backtracking with extrusion of RNA 3′-terminus through the secondary channel and EC inactivation (iv), or RNAP forward translocation which would bring the EC to an active post-translocated state (i). Image adapted from (Borukhov y Nudler, 2008)

The energy of NTP hydrolysis is not directly used for RNAP translocation along the DNA. The RNAP moves as the Brownian ratchet machine, driven forward by the binding of correct substrate (Vassylyev *et al.*, 2007a). This means that the RNAP thermally fluctuates between pre and post-translocated states of the substrate free EC, while the binding of the correct substrate would favor the post-translocated one (moves the equilibrium towards the post-translocated state).

Requirement for the presence of an additional protein defines the mode of transcription termination in bacteria. Thus, Rho-dependent and Rho-independent termination can be distinguished. The Rho-independent termination rests on the intrinsic terminator sequence, composed of a GC-rich palindrome sequence and an oligo-T sequence (the T-stretch). They are responsible for the synthesis of RNA molecules with a hairpin structure and the string of 7-8 uracil residues (the U-stretch). Remarkable stability of the EC, accounting for its high processivity, is based on RNAP´s interactions with the double-stranded DNA (double-stranded DNA-binding site (DBS)), the DNA-RNA hybrid (hybrid binding site (HBS)) and the single-stranded nascent RNA (RNA binding strand (RBS)) (Nudler *et al.*, 1998, von Hippel, 1998). Therefore, the net stability of the EC can be defined as the sum of three standard free energy changes, the DNA binding in DBS in (ΔG_{DBS}), the hybrid binding in HBS (ΔG_{HBS}), and the RNA binding in RBS (ΔG_{RBS}). Diminishing the contribution of only one of the components is not enough to permanently destabilize the EC, and to provoke its rupture and transcription termination. In the EC, the region between nts at positions 8 and 17 upstream of the RNA´s 3'end establishes contacts with at least 3 domains of the RNAP (RBS): β' (residues 29-58), β' (residues 298-330) and β (residues 1232-1304) (Gusarov y Nudler, 1999). At the same time, 8 nts of the nascent RNA form the hybrid in the transcription bubble (Nudler *et al.*, 1997). Upon the encounter with the T-stretch (distal part of T stretch, T7-T8), the RNAP has a tendency to backtrack, giving enough time for the hairpin to be formed. The hairpin formation occurs as the result of low A:U base pairing energy (proximal part of T-stretch, T1-T8) and the competition between the hybrid and the hairpin. The hairpin disrupts most of the A:U bps (the hybrid now only comprises 4 bps) and destroys most of the interactions at RBS (β' (29-58) and β (1232-1304), decreasing the contribution of ΔG_{HBS} and ΔG_{RBS}, at the same time (Gusarov y Nudler, 1999). The RNA would clash with the G- loop, resulting in irreversible "pausing" of the EC. Forced G- loop motion leads to opening of the clamp and release of the nucleic acids.

Rho-dependent terminators make almost the half of all terminators in *E.coli*. They are 70-90 bases long, consisting of Rho binding or utilization site (rut) and multiple transcription stop points (tsp) (Zhu y von Hippel, 1998). Rho factor, a hexameric protein, binds to a primary transcript rich in G and poor in C content. This binding leads to activation of its ATPase activity, which provides the protein with energy necessary for translocation along the transcript (in 5'-3' direction). The released energy is also used for the helicase activity, which peals of the primary transcript from the DNA-RNA hybrid (Ciampi, 2006).

1.3.2. Single-subunit RNA polymerases

Although single and multi-subunit RNAPs have different architecture, they share a large number of functional/biochemical characteristics. This, together with their relatively simple structure, made single subunit RNAPs the most suitable subject for studying the process of transcription, from the structural, structurally-functional and mechanistic point of view.

The T7 bacteriophage RNAP, a 98 kDa protein, is the most examined and the best characterized member of this group. As the rest of RNAPs, it exhibits three phases of transcription– initiation, elongation and termination, and two active states– initiation (abortive) mode and elongation (processive) mode.

1.3.2.1. Structural organization of T7 phage RNA polymerase

The X-ray crystal structure of T7 RNAP (Sousa *et al.*, 1993) and subsequent crystal structures of this protein in various complexes and states (Jeruzalmi y Steitz, 1998, Cheetham y Steitz, 1999, Yin y Steitz, 2002, Durniak *et al.*, 2008) have revealed that despite the limited sequence similarity, T7 RNAP displays fairly similar architecture to DNA polymerase I family (almost identical folding topology and secondary structure) (Sousa y Mukherjee, 2003). T7 RNAP has a form of cupped right hand and its subdomains, arranged around the binding cleft, has been named according to the similarity with the Klenow fragment:

- **The thumb domain**, including amino acids from the position 326 to 411, is a long α helical structure that protrudes from one side of template-binding cleft (Jeruzalmi y Steitz, 1998) (Figure 7). This structure is implicated in the stabilization of the transcription complex and in maintaining its processivity during the elongation (Sousa y Mukherjee, 2003).

- **The palm domain** comprises residues 412-553, and 785-883. It contains two β-hairpins, shown to be implicated in the initial steps of transcription and to be crucial for the catalytic activity of the enzyme (Jeruzalmi and Steitz 1998) (Figure 7). The β-hairpin formed by the residues 417-429, interacts with 0-2 bps of the DNA-RNA hybrid, playing a role in its stabilization at the early stages of transcription initiation (Cheetham y Steitz, 1999, Imburgio *et al.*, 2002). The amino acids from the position 805 to 817 form a β hairpin, with aspartic acids 812 (Asp 812) of the tip, implicated in coordination of catalytic Mg^{2+} ions.

- **The fingers domain** consists of residues 554-784 (Figure 7). This domain contains several helices responsible for the correct NTP selection and discrimination against deoxy NTPs (dNTPs) (Jeruzalmi and Steitz, 1998). Helices 29 (residues 626-632) and 30 (residues 636- 640) make an important interaction with the incoming rNTPs. K627 and K631 contribute to affinity of NTP binding through the interactions with phosphate groups, while Y639 is involved in NTP binding, and in the catalysis and selection of the rNTPs (Sousa and Mukherjee, 2003). It has been shown that this residue is conserved between RNAP and DNAP and that further discrimination toward dNTPs in DNAP is obtained by the presence of Glu residue in its close proximity that sterically

prevents binding of rNTP. Helices 30 and 31 (residues 642- 654), and the connecting turn establish most of the contacts between the template base at position i, the one that will be paired with incoming rNTP, and the very first base downstream of it (position i+1). These contacts are important for the fidelity of the transcription (Huang *et al.*, 2000, Huang y Sousa, 2000). Fingers domain also contains a β hairpin insertion that spreads from 739 to 770 and is termed specificity loop, due to its role in promoter recognition and binding (Temiakov *et al.*, 2000).

The structural comparison with the Klenow fragment has confirmed the sequence analysis data about the existence of an extra, novel N-terminal domain (residues 1-325) (Figure 7). This novel domain, entitled promoter binding domain (PBD), contains a six helix bundle (residues 72-150 and 191-267). The PDB together with the specificity loop of the fingers domain is essential for promoter recognition and binding, therefore for the transcription initiation (Muller *et al.*, 1988, Durniak *et al.*, 2008).

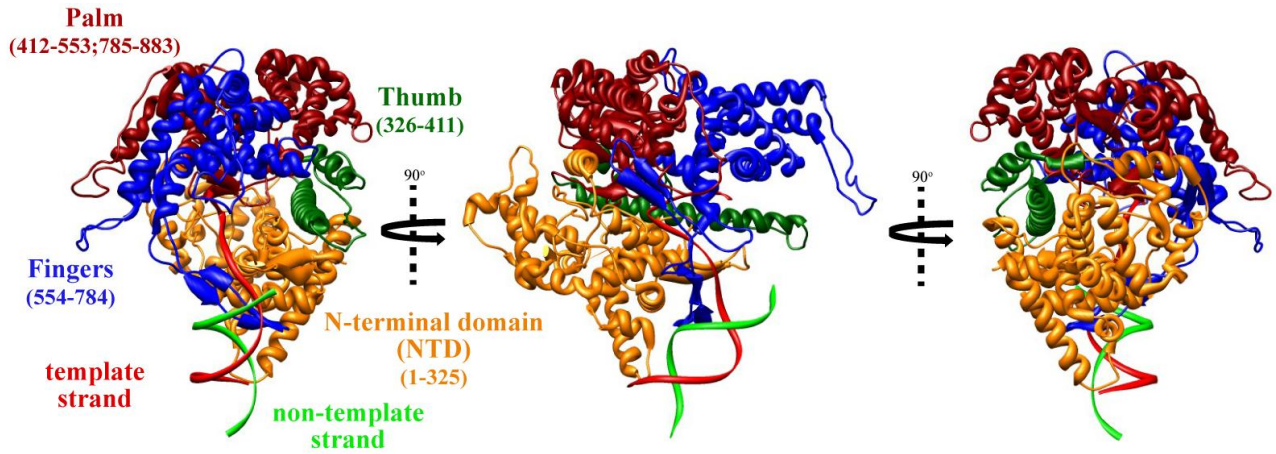


Figure 7: Structure of the T7 RNAP initiation complex. The X-ray structure of T7 RPo (PDB reference 1CEZ) in three typical views: frontal, side, and back (from left to right). The crystal structure resembling a cupped right hand is shown with its subdomains: NTD (orange); thumb domain (dark green); palm domain (dark red); fingers (dark blue); template and non-template strands of DNA are dyed in red and green, respectively.

1.3.2.2. T7 RNAP transcription cycle

Unlike the multi-subunit RNAPs, T7 RNA polymerase does not require the presence of any additional protein for the transcription initiation, which can be broken into several steps: (a) promoter recognition, binding and promoter melting; (b) initial synthesis of RNA up to 8 nts; and (c) promoter release and transition to stable elongation state (RNA transcript ~9-14 nts long), including the major conformational rearrangements, largely in the N-terminal portion of the enzyme.

The T7 bacteriophage genome contains 17 promoters, which can be functionally classified in 3 classes: (1) Replication promoters; (2) Class II promoters; and (3) Class III promoters. The Class II and III promoters are involved in expression of T7 phage's middle and late genes, respectively (Ikeda, 1992).The consensus sequence of the Class III T7 promoters is 23 bps long, spreading from positions -17 to +6 (Figure 8A).

It can be functionally divided into two parts:

- **Binding domain** (bps from-17 to -5) as its name states, is involved in recognition and binding by T7 RNAP. This region stays in duplex form during the whole initiation process.
- **Initiation domain** (bps from -4 to +5) is involved in melting and adequate positioning of +1 base of the template strand into the active site.

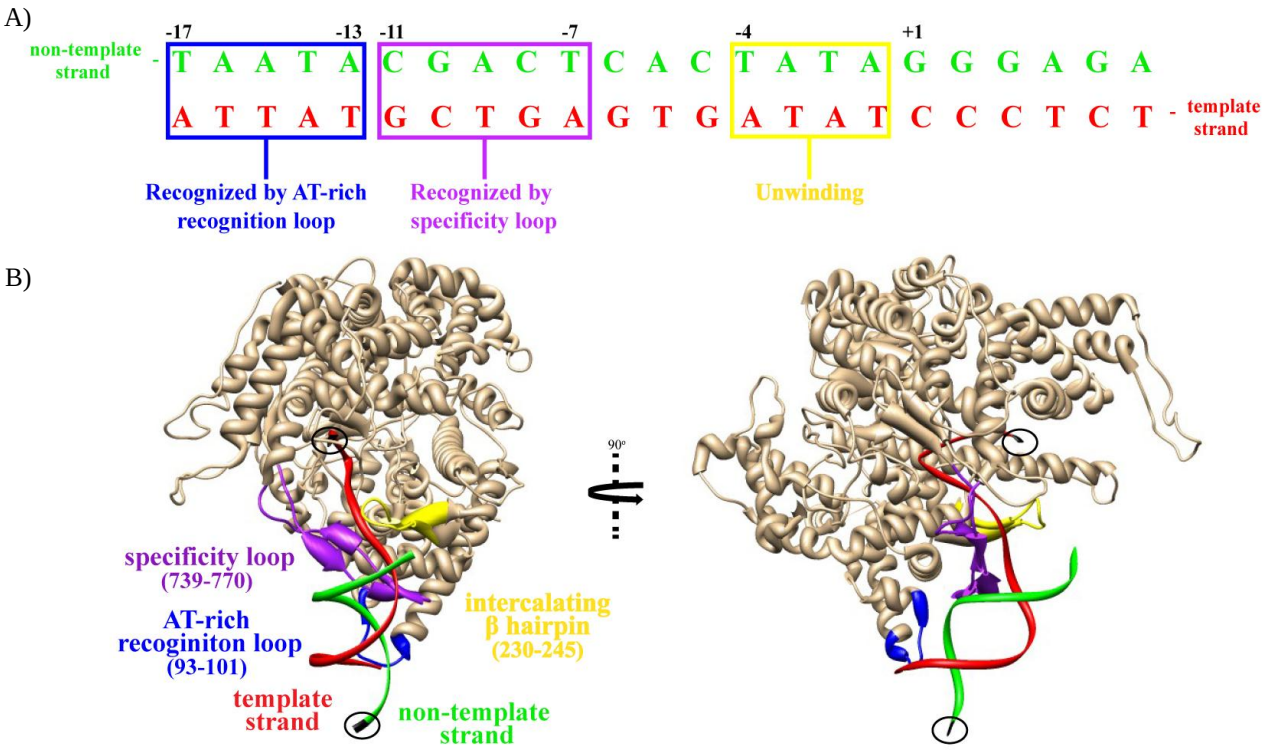


Figure 8: Promoter recognition elements of T7 RNAP and class III promoter. A) Two loops from N-terminal domain (AT rich recognition loop, blue, and intercalating β-hairpin, yellow) and one from fingers (specificity loop, purple) interact with the promoter. 5' termini of non-template and template DNA strands are encircled and colored in black. **B)** Class III T7 promoter. Frames stress the parts of the class III promoter interacting with 3 different elements of T7 RNAP. The same color code is used as for A).

Promoter binding involves three structural motifs of T7 RNAP (Figure 8B): (1) the AT rich recognition loop (residues 93- 101); (2) the specificity loop (residues 739- 770); and (3) the intercalating β-hairpin including Val237 (residues 230- 245). The AT rich recognition loop recognizes the DNA sequence from position -17 to -13 of promoter. The insertion of the AT rich recognition loop in the minor groove, results in the phosphodiester backbone disorder and formation of a wider (change from 5.7 Å to 7 Å) and shallower (change from 7.5 Å to 4.4 Å) minor groove. This provokes a slight bend in the axis of the DNA helix, making the region between 11 to 7 bp upstream of starting site visible and accessible to the specificity loop (Cheetham *et al.*, 1999). Contacts between the RNAP and the promoter are established at positions -11 and -10 of non-template strand, while from position -9 to -6/-5 they involve the template strand (Rong *et al.*, 1998). The specificity loop contains two residues, N748 and Q758, responsible for the recognition of the bases at the positions -11/-10 and -8, respectively. The mutational studies have shown that substitution of the N748 with the corresponding residue of T3 RNAP, changes the preference of the protein towards the T3 promoter (Temiakov *et al.*, 2000).

The initial binding of the RNAP to DNA induces a 40-60^{o} DNA bending around the -1 bp (Figure 9). This leads to accumulation of tension in the promoter, which is relieved by the opening of the double stranded DNA in the TATA region of the initiation domain (upstream edge of the promoter bubble). Also, the previous binding events allow the T7 RNAP to get properly orientated and positioned towards the DNA, placing the intercalating β- hairpin (residues 230- 245) in close proximity of the initiation domain of the promoter. The upstream part of the promoter bubble is maintained open through the insertion of the β-hairpin with the V237 in place of connection of bps -4 and -5 (Cheetham *et al.*, 1999, Cheetham y Steitz, 1999). V237 separates the template from the non-template strand and directs the template towards the catalytic active site. The 5`end of the template strand then gets to be bound to the pocket formed in the junction of the palm, the fingers and the N-terminal domains, through the base-stacking interaction with W422. The aromatic side chain of W422 stacks on the -1 base, inducing the sharp bend in the template strand and positioning the base at the position +1 as the initial one. Correct localization of the +1 base of the template strand into the active site and the presence of guanosine-5'-triphosphate (GTP), allows the incorporation of the first nucleotide. The enzyme-catalyzed reaction begins by placing the substrate, another molecule of GTP, at the position +2 and formation of phosophodiester bonds with GTP at the position +1. At that moment, the transcription bubble is 7 or 8 bases long, reaching the positions +3 or +4.

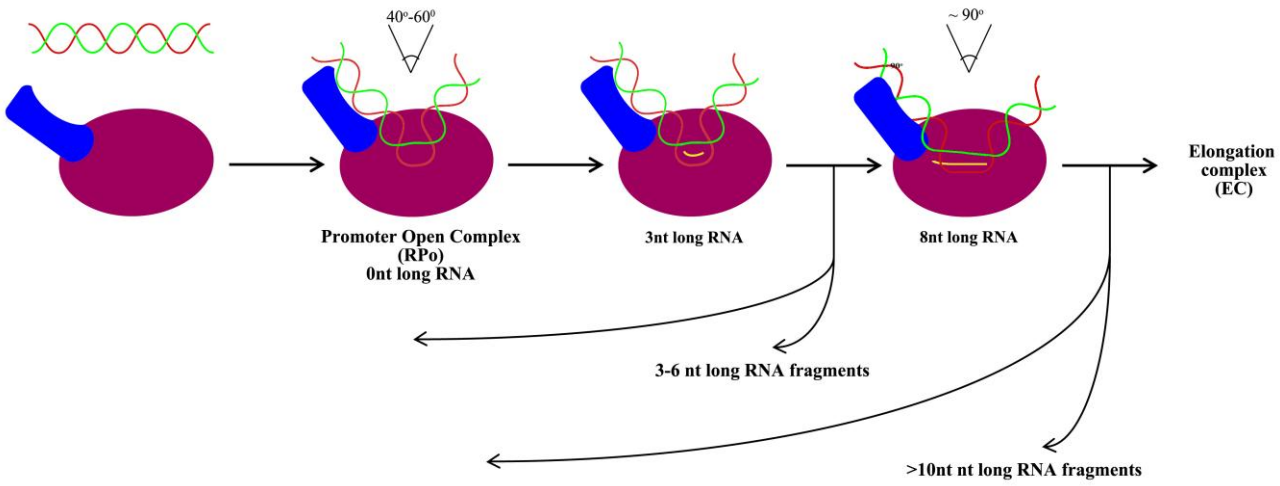


Figure 9: Schematic illustration of major T7 phage RNAP transition steps during the transcription cycle from core to the EC. Color legend: blue– N-terminal domain (NBD); dark red– C- terminal part (thumb, palm, fingers), red– template DNA strand; green– non-template DNA strand; yellow– RNA. The bending angle of DNA in RPo is 40-60^{o}. The left hand rotation of NTD for 45^{o}, upon the incorporation of the 8th nt in RNA molecule, provokes the additional DNA banding of 30-35^{o} (assuming the final bending angle of ~ 90^{o}).

The extension of the RNA beyond 3 nts is conditioned by the conformational changes of the enzyme, which widen the binding cleft and enable accommodation of the DNA:RNA hybrid longer than 3 bps (Cheetham y Steitz, 1999). On the other hand, the transcription bubble has to be expanded downstream, while it keeps the contacts with the RNAP, upstream. Retention of the upstream contacts is important for the stabilization of the complex, at the early steps, as the RNA is not long enough for more interactions to be established with the protein. This also means that the template DNA is pulled into the protein and scrunched as the RNA is extended. When the RNA reaches the length of 5 nts the thumb subdomain bends towards the RNA and the template strand, an important conformational change for the stability and processivity of the

transcription complex (Sousa y Mukherjee, 2003). These early steps of RNA synthesis probably involve several conformational intermediates, which due to the scrunched DNA (accumulated stress) and twisted subdomains are highly unstable. Therefore, there is a tendency of these high-energy intermediates to acquire more stable conformation, most probably returning to a start point, the initiation open complex. This results in the release of a 2-6 nts long RNA. When the length of the DNA-RNA hybrid reaches more than 7 bps, the base paring energy would be higher than the one needed for the dissociation of the hybrid, making the disruption of the RNAP-promoter contacts more probable.

The addition of first 8 bases in nascent RNA is followed by a left handed rotation of the PBD (residues 72-152 and 191-267 of N-terminal domain) and specificity loop for 45^{o} (Figure 9), which allows the growing DNA-RNA hybrid to be accommodated in the enlarged binding cleft (Steitz, 2009). This rigid body rotation does not affect the contacts with the promoter. As the DNA remains bound to the promoter, it is additionally bent for 30^{o} - 35^{o} (respective to the initial bending caused by the T7 RNA binding) and its upstream and downstream parts are brought to the distance of 6Å (Durniak *et al.*, 2008). As the RNA reaches the length of 8 nts, the transcription bubble spreads from the positions -4 to +9. This means that the transcription bubble is 13 or 14 bases long, as it is closed 1-2 nts downstream of RNA 3' end (Huang y Sousa, 2000).

Upon the incorporation of a base at the position +9 of RNA, the upstream edge of the transcription bubble collapses (Martin *et al.*, 2005). This is the result of accumulated tension of the scrunched DNA and competition of the RNA 3' end with the promoter for the specificity loop (Sousa y Mukherjee, 2003). Once the specificity loop is extended over its limits and its contacts with the promoter are disrupted, the intercalating β- hairpin should be displaced and its contacts with the DNA (at the connection of bps -4 and -5) should be lost (Martin *et al.*, 2005).

None of the conformational changes are defined, or how they occur, after the RNA transcript is extended beyond 8 nt. What is known is that the N- terminal region and the specificity loop undergoes large structural rearrangement resulting in the loss of all contacts with the promoter and formation of H sub domain, the major part of the RNA exit channel (Yin y Steitz, 2002). The C-terminal domain undergoes fewer conformational changes: (a) The loop at the tip of the fingers domain (residues 593- 612) moves towards the downstream DNA, forming the part of DNA binding site and (b) the N- terminal part of the thumb's α- helix bends towards the fingers domain, completing the formation of the pore that leads to the active site (Tahirov *et al.*, 2002).

The collapse of the upstream edge of the promoter from the position -4 to -1 facilitates the collapse at positions +1 and +2. The latter is important for subsequent release of the 5'end of the RNA, allowing it to go into the previously formed RNA exit channel (Martin *et al.*, 2005). The incorporation of a new NTP in the growing RNA molecule is a 4 stage process. Substrate binds to T7 RNAP in the open state. The NTP binding pocket is at the O helix and outside of active site. The bound NTP is base-paired to the next template base, but it is not in the insertion position. The binding of NTP induces the transition of the enzyme to the closed

state, and the correct placing of the substrate in the insertion position. This state is stabilized by ionic interactions between the three phosphate groups of recently bound NTP and the 2 Mg^{2+} ions bound to the T7 RNAP on one side and the arginine (R) of O helix on the other. T7 RNAP remains in the same closed conformation even after the nucleotide has been incorporated into the RNA. This time the closed state is held by ionic crosslink formed with PPi, produced during the incorporation of nucleotide, and Mg^{2+} ion of the active site and the R on the O helix. Upon the release of the PPi, the T7 RNAP exhibits rotation of the O helix and returns to the open state. This conformational change also moves the Y639 to the newly formed primer terminus, preventing the binding of a new NTP to the insertion site and disabling transition to the closed state. The binding of the next NTP to the preinsertion site displaces the Y639 from the insertion site (Steitz, 2009).

1.4. Mitochondrial transcription machinery

The mitochondria can be observed as the power plants of the eukaryotic cells, due to their main function, the synthesis of adenosine triphosphate (ATP). Mitochondria, however, play additional important roles in a plethora of cellular processes such as cell growth, cell differentiation, programmed cell death (apoptosis), signaling and innate immunity (McWhirter *et al.*, 2005, Rehman, 2010, Mammucari *et al.*, 2011, Martinou y Youle, 2011, Cloonan y Choi, 2012). Therefore, they are an essential, indispensible part of every living eukaryotic cell.

In the 1962, the clear link between mitochondrial malfunction and human diseases had been reported (Luft *et al.*, 1962). Today, the existence of different mitochondrial disorders is well established (Wallace, 1992). Mitochondrial dysfunctions are in the base of broad range of human diseases, mostly affecting heart, skeletal muscles and central nervous system. Due to the fact that mitochondria are semi-autonomous organelles, mutations in either of genomes, mitochondrial and nucleus, could lead to their dysfunction. Pathogenic mitochondrial DNA mutations were first reported in 1988 (Holt *et al.*, 1988, Wallace *et al.*, 1992), and evidence that nuclear genes could also be involved in mitochondrial disorders, was published almost two years later (Zeviani *et al.*, 1990). All this is the origin of the strong desire of scientist for deciphering the functions and structure of mitochondria, to their tiny bits and pieces.

1.4.1. Origin of mitochondria and organization of mitochondrial DNA

The most acceptable explanation for mitochondrial origin is provided by the endosymbiotic theory, which has been postulated in the 19th century. Nevertheless, the broad approvement of this theory came with the appearance of molecular methods that allowed some of its predications to be tested (Margulis, 1970). According to the endosymbiotic theory, mitochondria were created 2 billion years ago through the invasion of archaeon or primitive eukaryote by the ancestral α-proteobacterium, also a common ancestor of Rickettsia, genus of obligatory intercellular parasites (Andersson *et al.*, 1998). This most likely unique event in course of evolution was preceded by a drastic change in concentration of atmospheric oxygen, from 1% to more

than 15%, which happened over a relatively short time span (200 million years). The ancestral α-proteobacterium was an aerobic organism capable of oxygen consumption with the aid of a respiratory chain, ending with cytochrome oxidase. On the other hand, the heterotrophic, anaerobic host provided the endosymbiont with pyruvate, a starting precursor of the Krebs cycle. It is thought that at the increasing levels of oxygen and in the lack of widespread protective mechanisms such as peroxidases, catalases, and superoxid dismutases, the consumption of oxygen by cytochrome oxidase would be a significant evolutionary advantage (Kurland y Andersson, 2000). Over the time, the endosymbiont genome underwent huge reduction and degradation. In fact, the gradual loss of genes was the major driving force of endosymbiont-organelle transition (Lithgow y Schneider, 2010). This is the result of synergistic activity of several mechanisms: (a) asexual reproduction, making the obligate symbiont more prone to accumulation of sublethal mutations, including inactivation and loss of non-essential genes (heavier mutational load) (Gray, 1989); (b) rise in the number of non-essential genes once the symbiotic relationship is established, due to the new environment abundant with host's metabolic intermediates; (c) continues gene transfer. Besides, it has been shown that asexual genomes with small population size are affected by the Muller's ratchet, the evolutionary mechanism which explains the accumulation of irreversible deleterious mutations in asexual populations (Bergstrom y Pritchard, 1998). As the consequence, the mutation frequency per genome is proportional to the size of the genome, implying that larger genomes are more vulnerable, as it was the case in the early steps of mitochondrial evolution.

Experiments on *S. cerevisiae* have revealed not just the existence of a continuous gene transfer from the mitochondria to the nucleus, but also the mechanism by which this is committed (Lithgow y Schneider, 2010). The mitochondrial DNA fragments are released in the cytosol through the process of autophagy, performed by the phagolisosomes (Thorsness y Fox, 1990, Thorsness y Fox, 1993). The liberated nucleic acid fragments can become intermediates for the transfer in nucleus. The rates for wild type (wt) *S.cerevisae* are estimated to be once in 10^5 generations (Thorsness y Fox, 1990). The heavier mutational load and the existence of a transfer mechanism favored the transfer of mitochondrial genes to the nucleus. These nuclear copies of mtDNA are named the "numts". For example, the yeast nuclear genome contains 34 numts that range in size from 22 to 230 bp and are 80–100 % similar to mtDNA (Timmis *et al.*, 2004). The inactive genes were transferred while the active versions remained in mitochondrial genome. Once the inactive genes gained the tagging, so the protein products could be transported back to mitochondria, the redundant mitochondrial genes were eliminated through the mutational degradation. According to this, the reason for the expansion of splicing system among the primitive eukaryotic genomes is to satisfy the need for tagging of newly transferred genes from mitochondrion and chloroplasts (Lithgow y Schneider, 2010). The genes, whose transfer wasn't neutral, were maintained in mitochondria. The ancestral genes that had the homologues in the host genome were replaced with the homologous genes, leaving only the small fraction of original endosymbiotic genes. In parallel, the compensation was followed by the addition of new eukaryotic gene products to the mitochondrial proteome. Functional genomics studies show that almost the half of

mitochondrial proteome, cca. 200 genes, in *S. cervisiae* consists of novel recruits from the eukaryotic genomes (Lithgow y Schneider, 2010).

In yeasts, mtDNA is 76kb long and with abundant non–coding regions, introns and promoters. It encodes components of the mitochondrial translation machinery (21S and 14S rRNAs, 24 tRNAs, RNA-subunit of mtRNase P, and one ribosomal protein) and important proteins for ATP synthesis. mtDNA is not distributed homogeneously throughout the matrix, rather concentrated in structures called nucleoids. Nucleoids are associated with the inner mitochondrial membrane, but this association and their function are not fully understood. It is thought that they might play roles in inheritance, segregation, replication, transcription and probably recombination. Their abundance in yeasts depends on the growth conditions and genetic background (Lockshon *et al.*, 1995, MacAlpine *et al.*, 2000). At least 20 protein components were identified in yeast nucleoids, among them: abundant high-mobility group DNA binding protein (Abf2), aconitase, alpha-keto-beta-hydroxylacil reductoisomerase (Ilv5), mitochondrial distribution and morphology proteins 10 and 12 (Mdm10p and Mdm12p).

1.4.2. Structure of mitochondrial transcription machinery

Based on the mitochondrial origin one could expect the employment of a complex transcription machinery in expression of mitochondrial encoded genes. Mitochondrial transcription machinery, however, is distinct from the nuclear and cannot be assigned to the multi-subunit RNAP family. It is a much simpler system which consists of: (a) a single unit core RNA polymerase (mitochondrial RNAP or mtRNAP), and (b) one (yeast) or two additional accessory molecules (mammals). Much of our knowledge and understanding of how the mitochondrial transcription is performed and regulated, including in humans as well, rests on intensive studies on *S. cervisiae* transcription. Yeasts, in general, are commonly used models of eukaryotic organisms in molecular biology studies, because the basic mechanisms of DNA and chromosomal replication, cell division, gene expression, translation, signal transduction and regulatory networks, metabolism, and sub-cellular organization are essentially conserved between yeast and higher eukaryotes (Castrillo y Oliver, 2004, Balaji *et al.*, 2006, Mustacchi *et al.*, 2006). Furthermore, yeasts are free-living organisms with rapid growth, simple methods of cultivation under defined conditions, easy replica plating, and mutant isolation all that makes them suitable candidate for biological studies. They are also well-defined genetic system with simple techniques of genetic manipulation and the first eukaryotic organism with completely sequenced genome (Goffeau *et al.*, 1996).

In yeast, mitochondrial transcription is performed by a two component system, composed of RPO41 gene product (core RNA polymerase, sc-mtRNAP), and the accessory transcription factor, mitochondrial transcription factor 1 (Mtf1). This machinery can start transcription from at least 20 different starting sites, which contain well conserved consensus nonannucleotide sequence ($^{-8}$ATATAAGTA^{+1}). Base content at the positions +2 and +3 defines the strength of promoter, thus being responsible for the different expression rates of different yeast mitochondrial genes. Promoters with purine at +2 and pyrimidine at +3 position in the non-

template strand are strong promoters, while the weak ones are promoters with pyrimidine and purine at +2 and +3 positions, respectively. Genes with the strong promoters have 15-20 times higher levels of expression than the genes with the weak promoters.

Sc-mtRNAP, 1351 amino acids long and with molecular mass of 153 kDa, contains 8 regions of homology with T3/T7 phage RNAP, thus often comprised under single-subunit RNA polymerase family, together with all mtRNAPs. Those regions consist of 9-89 amino acids, with identity of 49%-77%. In total, these two proteins "share" 219 amino acids and they are 28% identical (Masters *et al.*, 1987). The existence of evident sequence similarity between Sc-mtRNAP and T3/T7 phage RNAPs led to the assumption that the homology goes beyond just a sequence level. It has also raised several questions about the origin of transcription in general, which of two RNAP families appeared first, and where did the ancestral single-subunit RNAP gene come from. The discovery of eubacteria-like multi-subunit RNAP in *Reclinomonas americana*'s mitochondrial transcription has given a support to a view that that mitochondrial system initially employed a multi-subunit RNAP, that was along the course of evolution replaced by the nucleus encoded single-subunit RNAP. The phylogenic studies showed that this "exchange" has occurred early in the evolution of eukaryotic lineage, shortly after the mitochondrial endosymbiosis (Cermakian *et al.*, 1996). However, it is still uncertain whether the single-subunit RNAP gene comes from T7-phage or plasmid-like DNA, or that those two are derived from the nuclear gene. On the other hand, it seems unlikely that the single-subunit RNA gene was already present in the host nuclear genome before the endosymbiosis or that it was introduced by eubacteria-like endosymbiont.

Recently published atomic structure of human mtRNAP confirmed the assumption regarding its similarity to T3/T7 phage RNAP at structural level (Ringel *et al.*, 2011). Based on the amino acid sequence similarity mtRNAP can be divided into (Figure 10A-C):

- **C-terminal domain**, where most of the homology boxes are located (residues 648-1230), including the elements responsible for establishing interactions with the template DNA strand, NTP and Mg^{2+}, as well as the catalytic site. This implies the common mechanism of RNA synthesis (incorporation of NTPs in nascent RNA), that is shared among T7 RNAP and mtRNAPs. The specificity loop, T7 RNAP´s element implicated in promoter recognition and melting, was identified in sc-mtRNAP (residue 1127-1149) and human mtRNAP (residues 1086-1105) by sequence analysis. Furthermore, mutational and chemical-nuclease studies have shown the existence of interaction between specificity loop and region between -8 and -1 of the yeast promoter, identifying E1224, K1127, Q1129, Q1135, and T1136 as the crucial amino acids (Matsunaga y Jaehning, 2004, Nayak *et al.*, 2009). The large part of the specificity loop is not visible in the crystal structure of human mtRNAP, indicating that this element assumes defined structure through the interaction with the promoter.
- **N-terminal domain** (residues 369- 647) possesses moderate sequence similarity with the NBD of T7 RNAP, enclosing the AT rich recognition loop (sc-mtRNAP: residues 481-497; human

mtRNAP: residues 450-470) and the intercalating β hairpin (sc-mtRNAP: residues 617-630; human mtRNAP: residues 610-620) (Nayak *et al.*, 2009, Ringel *et al.*, 2011). Mutational studies of five amino acids of the specificity loop tip have showed the importance of this element in promoter melting, affecting the transcription from the double-stranded but not from the pre-melted promoter. The function of the AT rich recognition loop in mtRNAPs, however, is not well defined, as the mutations failed to provoke any change in human mitochondrial transcription (Ringel *et al.*, 2011).

- **N-terminal extension** comprises approximately first 300 hundred amino acids. This is not just an extra part comparing to T7 RNAP, but also the most diverse part of the protein in all eukaryotic species. There are several assumptions and indications about the possible role of this N-terminal portion of the protein. Firstly, the first 42 amino acids of human mtRNAP make the mitochondrial targeting sequence (MTG). Secondly, there are results indicating the role of the first 185 amino acids in transcription–translation coupling, as the interactions with proteins involved in processes downstream of transcription have been shown (Rodeheffer *et al.*, 2001, Bryan *et al.*, 2002, Rodeheffer y Shadel, 2003, Markov *et al.*, 2009). Thirdly, deletion studies of the first 270 residues of sc-mtRNAP have led to conclusion of auto-inhibitory role of this portion, as the ability of protein to recognize and to melt promoter remains intact, accompanied with a syntheses of lower number of abortive products (comparing to the wild type (wt) protein) (Paratkar *et al.*, 2011). Furthermore, deletion of the first 380 amino acids incapacitates sc-mtRNAP to start the transcription from the double-stranded promoter, but not from the pre-melted one. Also, the interaction with Mtf1 is strongly affected, implying the role of the residues 270-380 in the interaction with Mtf1.

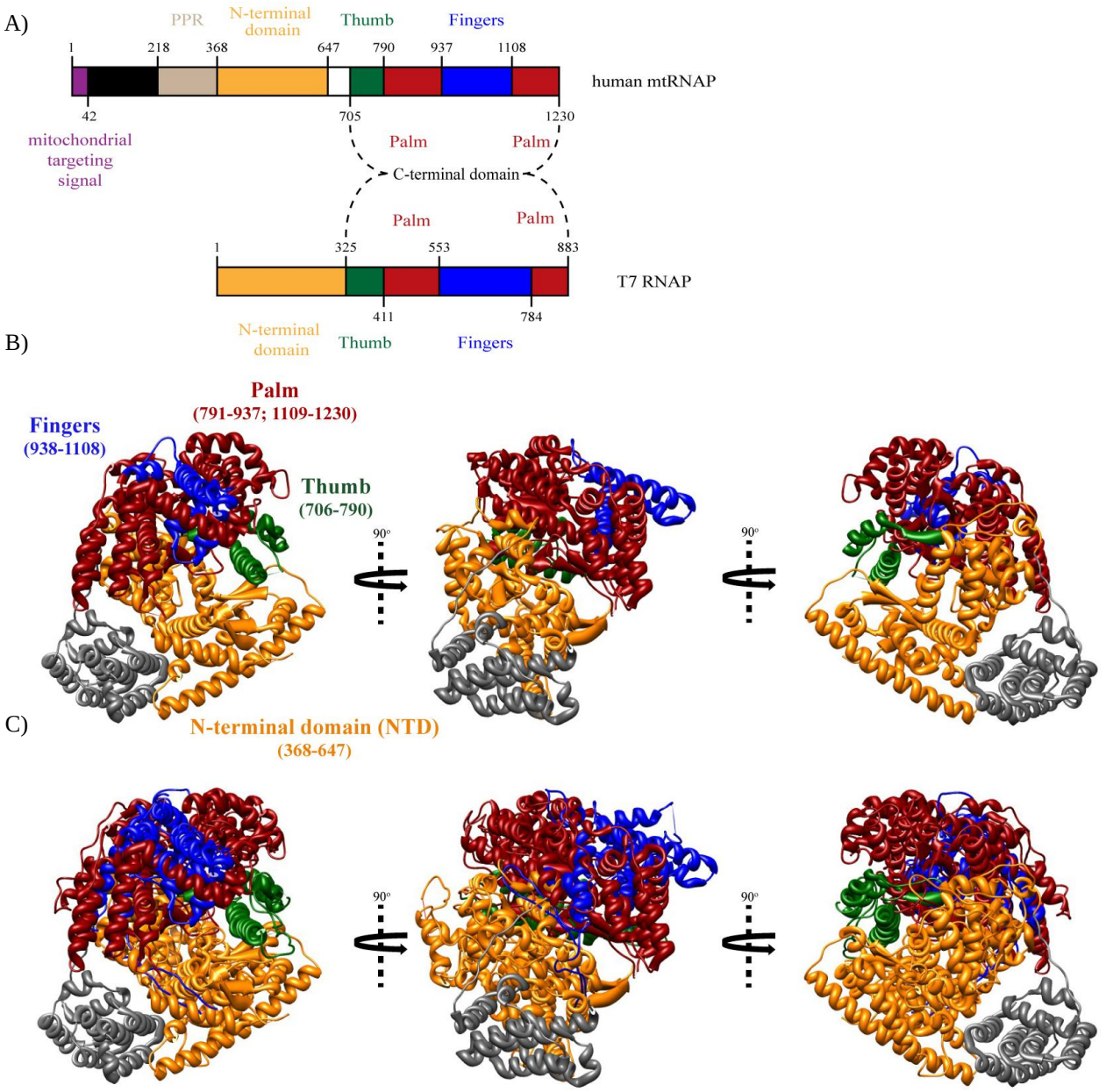


Figure 10: Structures of human mtRNAP and Mtf1. A) Schematic comparison of mtRNAP´s and T7 RNAP´s sequences. Regions exhibiting sequence homology with T7 RNAP are emphasized: NTD (orange); Thumb domain (dark green); Palm domain (dark red); Fingers (blue); mtRNAPs contain an additional N-terminal extension which is composed of MTG (residues 1-42, purple); region involved in transcription-translation coupling (42-218, black) and putative pentatricopeptide repeat (PPR), found in proteins implicated in RNA-processing in mitochondria and chloroplasts. **B)** X-ray structure of human mtRNAP (PDB 3SPA) is shown in three typical views: frontal, side and back. The same color code is used as for A). The template and non-template strands of DNA are dyed in red and green, respectively **C)** Structural comparison of human mtRNAP and T7 RNAP (PDB 3SPA and 4RNAP) at the same activity state.

Unlike the T7RNAP, mtRNAPs require the presence of accessory molecules, mitochondrial transcription factors, for the promoter specific initiation. A need for Mtf1 or *S. cerevisiae* mitochondrial transcription factor B (sc-mtfB) for mtDNA's maintenance and discovery of regions with a limited similarity to the conserved region 2 of bacterial σ factors led to the conclusion that Mtf1 resembles the *E.coli*'s σ factors (Jang y Jaehning, 1991). Later its X-ray atomic structure, however, surprisingly has shown that Mtf1 is homologous to *E. coli*'s rRNA small subunit methyltransferase A (KsgA) (Schubot *et al.*, 2001, Shutt y Gray, 2006). The dual function for one of two human homologues of Mtf1, h-mtTFB1, has been confirmed.

h-mtTFB1 is able to methylate *E. coli* small subunit ribosomal RNA at the same sites as KsgA (McCulloch *et al.*, 2002). Phylogenetic analyses strongly indicate that mtTFB arose from a DNA methyltransferase (DMT) gene of proto-mitochondrial endosymbiont, which at some point of evolution acquired a secondary function. A fact that both the fungal and metazoan mtTFB homologues exhibit transcription factor activity, indicates that this new function was acquired before the fungal-animal divergence. Mtf1 is a compact protein composed of: a) a N- terminal α/β domain, containing the S-adenosyl-L-methionin binding site; b) a C-terminal helical domain; and c) a C- terminal tail, with a very important role in promoter melting and transcription initiation as it interacts with melted template strand in the open complex (Deshpande y Patel, 2012) (Figure 11A). The C-terminal tail cannot be seen in the crystal structure and it is assumed that it gets arranged through interaction with the DNA, as in the case of the specificity loop of human mtRNAP. The basic cleft, thought to facilitate the DNA binding, is placed between the N- and C-terminal domains (Shadel y Clayton, 1995) (Figure 11B). It was shown that Mtf1 plays several roles in transcription initiation, from discrimination of the promoter from non-promoter sequences, through additional bending of the bound DNA, over formation and stabilization of the RPo, by interacting with nts at the positions -10/-9 and -9/-8 of the template strand (in double-stranded form) and melted nts at -5/-4 and +1/+2 of the non-template strand, to selection of a startpoint (Amiott y Jaehning, 2006, Paratkar y Patel, 2010).

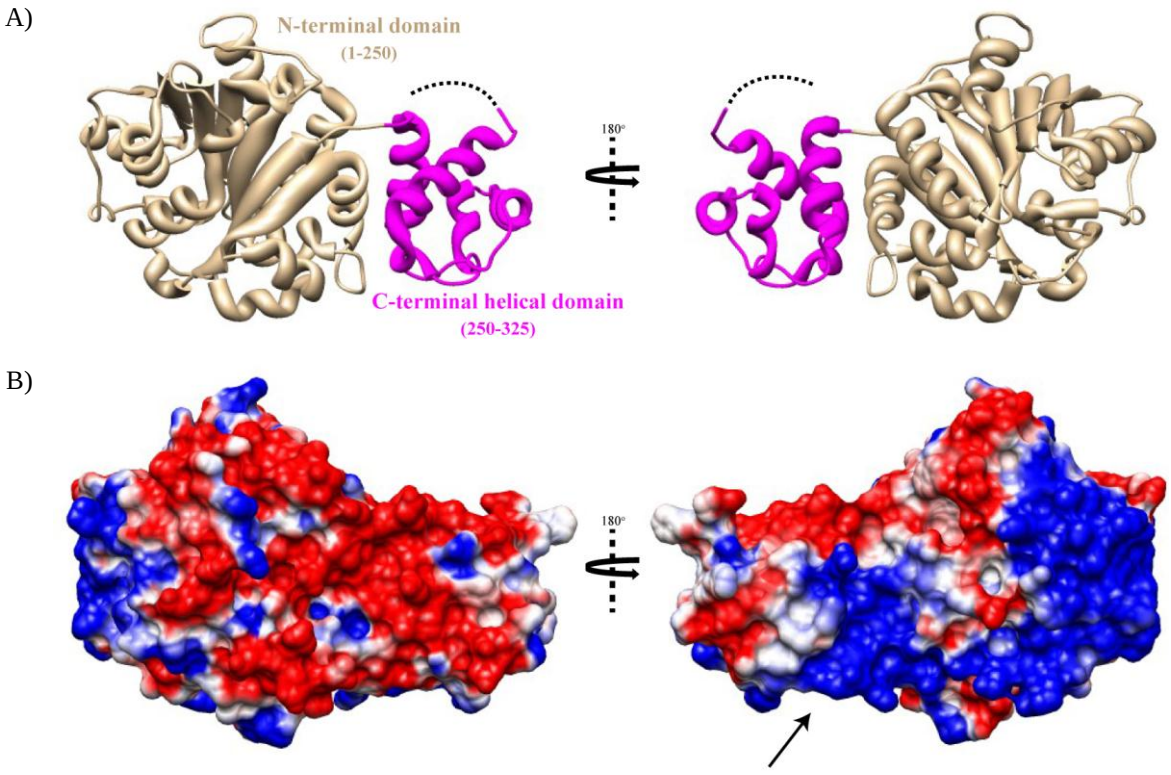


Figure 11: The structure of Mtf1. A) X ray structure of Mtf1 (PDB 1I4W), with N- (tan) and C- (magenta) terminal domains, is shown in two orientations. The dash curves represent the C terminal tail that is not solved. **B)** Distribution of electrostatic charges is also shown, emphasizing the existence of positively charged cleft (pointed with arrow), that might have a role in interaction with the DNA.

1.4.3. Importance of a more extensive structural study of the mitochondrial transcription machinery

The full comprehension of RNA synthesis, as the pivotal step in the expression of genetic information, yearns for structural, structural-functional and mechanistic studies of complex molecular machines, the RNAPs, that are at the base of this process. In the past few decades emerging structural data of bacterial, archaeal, nuclear and T7 phage RNAPs has contributed to understanding how the multi and single-subunit RNAPs work. Unfortunately, there is almost no structural information about mtRNAPs. The mitochondrial transcription has its roots in the interaction of three entities: the host cell, the bacteria and phage, thus, sharing a number of features with their transcription systems. Even though mtRNAPs are often placed under single-subunit RNAP family, their activity is dependent and regulated by a number of different factors and cofactors. On the other hand, once the transcription enters the elongation phase the mtRNAPs continue working as a single-subunit enzymes. This places the mtRNAPs in between two RNAP families. Therefore, the structural studies of mitochondrial transcription machinery would help in answering the questions regarding the origin and the evolution of transcription, and could also shed a light on the origin of the cellular life, as it is.

The clear link between some human pathologies and the mitochondria has been recognized. As the central place of energy production in cells, mitochondrial dysfunction mostly affects the energy consuming tissues and organs. The estimated frequency of inherited diseases caused by mutations in either nuclear or mitochondrial genes responsible for ATP production is once in every 4000 individuals, but they are difficult to diagnose due to broad clinical picture and absence of reliable screening methods. The age-related pathologies, however, are connected with normal mitochondrial activity, as the side products of ATP production are reactive oxygen species (ROS) that damage both cellular and organelle components. Firstly, it was believed that ROS-induced damage to mtDNA was the major cause of these pathologies, however, it has recently been found that mtRNA transcripts may be even more prone to oxidative damage (Li *et al.*, 2006). This suggests that balancing synthesis and degradation of mt transcripts is required to preserve the quality of mtRNAs and organelle integrity (Rogowska *et al.*, 2006). Therefore, structural studies of mitochondrial machinery would contribute to a better understanding of mitochondrial transcription control and its role in these pathological processes.

A unique X-ray structure of mtRNAP does not provide enough information about the interactions of the enzyme with either the nucleic acids or transcription factor(s). Some of the essential elements of T7 RNAP responsible for promoter recognition and melting, which presence has also been confirmed in mtRNAPs, are undefined in this structure. Thus, the structure of human mtRNAP can only be partially used to complement and verify the fidelity of existing models that explain the possible mechanism of activity. On the other hand, certain dilemmas regarding the interaction of mtRNAP with transcription factor cannot be solved. In order to give an answer to some of these questions, there is a need for the structural study of mtRNAP in various activity states.

2. Objectives

The general aim of this thesis was to fill the existing information gap regarding the structure of sc-mtRNAP, which would help in deciphering the mechanisms that are in the basics of the yeast mitochondrial transcription. In particular the following objectives have been pursued:

1. Three dimensional reconstruction of quasi-IC and DNA localization within the complex
2. Three dimensional reconstruction of IC and identification of the DNA's termini and possible interaction sites between the mtRNAP and Mtf1.
3. Three dimensional reconstruction of EC and determination of the DNA position.
4. Define the position of the N-terminal domain by three dimensional reconstruction of N- terminal deletion mutants and identification of the DNA within the complexes

3. Materials and Methods

3.1. Biological material

3.1.1. Protein samples

The protein complexes analyzed in this thesis were provided by Prof. Rui Sousa (Department of Biochemistry, The University of Texas Health Science Center at San Antonio, Texas, USA). The expression and purification of sc-mtRNAP and Mtf1 was carried out according to previously described protocol (Matsunaga *et al.*, 2004). The transformation and over-expression were performed in *E. coli* because the earlier over-expression protocols in yeast included a multi-step process that didn't provide satisfying results; the contamination with Mtf1 was always present. This provoked serious problems when working with complexes that should not contain Mtf1. In brief, the transformed *E. coli* BL21 DE3 codon plus (RIL) (*Stratagene*) cells were inoculated in LB medium (1:50 ratio) that contained 100 μg/ml ampicilin and incubated with constant shaking of 250 rotations per minute (rpm) for 7h at 37°C or until the optical density of the growing cell culture (A_{600nm}) reached the value of 0.6. The expression was induced by the addition of 0.6 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG) for 4h 30 min. The cells were harvested by 20 min centrifugation at 5000 rpm at 4°C. The pellet was resuspended in lysis buffer (40 mM Tris-HCl, pH 7.9, 300 mM NaCl, 0.1% Tween 20, 1 mM phenylmethanesulfonylfluoride (PMSF), 15% glycerol, 1 mM ethylenediaminetetraacetic acid (EDTA), 1 mM dithiothreitol (DTT), Roche-Complete™ Protease Inhibitor Mixture Tablet) and mixed with lysozime (1mg/ml). The lysed cells were centrifuged for 30 min at 13000 rmp at 4°C, to separate the cellular debris from the soluble proteins. The supernatant was subjected to polyethyleneimine (PEI) precipitation (0.5 M NaCl and 1/20 of the lysate volume of 10% PEI, pH 7.9) followed by another centrifugation. The supernatant was saturated with 50% ammonium sulfate and subjected to another round of 30 min centrifugation at 16000 *rpm* and 4°C. The obtained pellet was dialyzed for 2h at 4°C against the dialysis buffer (40 mM Tris-HCl, pH 7.9, 150 mM NaCl, 0.1% Tween 20, 15% glycerol, 1 mM EDTA, 1 mM DTT and 20mM imidazole). The solution was applied to a Ni Sepharose 6 column (*GE-Amersham Biosciences*) previously equilibrated with 10 volumes of the dialysis buffer. Once the sample passed through the column, the absorption at 280nm returned to 0, the bound His-tagged proteins were eluted with 0.3 M imidazole. The fractions were pooled and loaded on heparin-Sepharose CL-6B resin (*GE Healthcare*) equilibrated with the dialysis buffer without imidazole. The protein was eluted using a linear gradient of 150 mM to 1 M NaCl in the dialysis buffer. The peak fractions were analyzed for purity by SDS-PAGE and the pure fractions were pooled and dialyzed against the sample storage buffer A (20 mM Tris-HCl pH 8.0, 0.5 M NaCl; 1 mM EDTA; 1 mM DTT; 50% glycerol (v/v), in H_20). The proteins were concentrated in an Amicon concentrator using a regenerated cellulose ultrafiltration membrane (MWCO-30kDa, *Millipore*). The concentration of the samples was measured employing the Bradford protein assay (Bradford, 1976). The purified proteins were aliquoted and stored at –80 °C. The complexes (Table 1) were formed by simple mixing of appropriate proteins with nucleic acids in a slight molar excess (1.5 times) in order to diminish the likelihood of aggregate formation. The incubation of 90 minutes to 2h at room temperature (RT) was followed by dialysis against the sample storage buffer B (20 mM Tris-HCl pH 8.0,

0.05 M NaCl; 1 mM EDTA; 1 mM DTT; 50% glycerol (v/v), in H_20). The biotin-streptavidin labeled complexes were formed in our lab. The streptavdin was purchased from *Sigma-Aldrich* (reference number 85878).

Table 1: List of analyzed protein complexes in this study

Name	Composition
Quasi IC	Sc-mtRNAP:IC-DNA
5' non-template labeled quasi IC (upstream labeled quasi IC)	Sc-mtRNAP:IC-DNA (containing 5' streptavidin-biotin labeled non-template strand)
5' template labeled quasi IC (downstream labeled quasi IC)	Sc-mtRNAP:IC-DNA (containing 5' streptavidin-biotin labeled template strand)
IC	Sc-mtRNAP:IC-DNA:Mtf1
5' non-template labeled IC (upstream labeled IC)	Sc-mtRNAP:IC-DNA:Mtf1 (containing 5' streptavidin-biotin labeled non-template strand)
5' template labeled IC (downstream labeled IC)	Sc-mtRNAP:IC-DNA:Mtf1 (containing 5' streptavidin-biotin labeled template strand)
EC	Sc-mtRNAP:DNA:RNA
5' non-template labeled EC (upstream labeled EC)	Sc-mtRNAP:DNA:RNA (containing 5' streptavidin-biotin labeled non-template strand)
5' template labeled EC (downstream labeled EC)	Sc-mtRNAP:DNA:RNA (containing 5' streptavidin-biotin labeled template strand)
single-peptide RNAP IV EC (spRNAPIV EC-sc-mtRNAP lacking 368 amino acids from its N-terminal portion)	spRNAPIV:DNA:RNA
5' template labeled spRNAPIV EC (upstream spRNAPIV EC)	spRNAPIV:DNA:RNA (containing 5' streptavidin-biotion labeled template strand)

3.1.2. Nucleic acids

During the oligonucleotides design, several requirements were observed. Firstly, the oligonucleotides had to exhibit partial complementarity, so the resulting dsDNAs could contain unpaired nts that would comprise the artificially premelted promoter (bubble promoter) in IC-DNA and EC-DNA (underlined sequences, Table 2). This was especially important for quasi-IC formation as it was previously shown that mtRNAP could only be transiently bound to dsDNA (Deshpande y Patel, 2012). In the case of IC-DNA, the oligonucleotides had to contain conserved nonanucleotide mitochondrial promoter sequence (sequence in red, Table 2) that would provide the binding of mtRNAP:Mtf1 complex. Secondly, their length had to be chosen in such a manner that the binding of more than one mtRNAP or mtRNAP:Mtf1 complex per one molecule of DNA would be prevented. At the same time, they had to be long enough to allow undisturbed binding of streptavidin to biotin labeled DNA's termini.

Table 2: List of nucleic acids used in this study

Name	Sequence	Complex
IC-DNA template strand	$^{5'}$TAAAACTATTCTTTATTACTTATATCGCG$^{3'}$	quasi-IC and IC
5' biotin labeled IC-DNA template strand	$^{5'}$(bio)TAAAACTATTCTTTATTACTTATATCGCG$^{3'}$	quasi-IC and IC
IC-DNA non-template strand	$^{5'}$CGCGATATGCAGCTTAAAGAATAGTTTTA$^{3'}$	quasi-IC and IC
5' biotin labeled IC-DNA non-template strand	$^{5'}$(bio)CGCGATATGCAGCTTAAAGAATAGTTTTA$^{3'}$	quasi-IC and IC
EC-DNA template strand	$^{5'}$GCCCTGCGCATGTCTTCCTCC$^{3'}$	EC
5' biotin labeled EC-DNA template strand	$^{5'}$(bio)GCCCTGCGCATGTCTTCCTCC^{3}	EC
EC-DNA non-template strand	$^{5'}$ACACGTTACGTTCGCCACGGC$^{3'}$	EC
5' biotin labeled EC-DNA non-template strand	$^{5'}$(bio)ACACGTTACGTTCGCCACGGC$^{3'}$	EC
EC-RNA	$^{5'}$GGAGGAAGAC$^{3'}$	EC

3.2. Protein analysis

3.2.1. Electrophoresis in denaturing polyacrilamide gels

The purity and stability of the samples were examined by denaturing polyacrilamide electrophoresis (SDS-PAGE). This is a commonly used method that separates proteins according to their electrophoretic mobility, which depends on their length and their charge. The acrilamide:bisacrilamide composition is selected in accordance with the molecular mass of analyzed protein samples, which, in the case of this thesis, was 10–12%. The samples were mixed with 6x SDS loading buffer (50 mM Tris-HCl pH 7.6; 2mM EDTA; 2% (w/v) SDS; 25% glycerol; 0.01% (w/v) bromophenol blue; 5% (v/v) β-mercaptoethanol, in H_20) to a final concentration of 1x, and heated at 95^{o}C for 2 min. Electrophoresis was run using the SDS electrophoresis buffer (25 mM Tris pH 8.3; 0.1% (w/v) SDS; 250 mM glycine) and at electric current of approximately 20 mA. The applied voltage of 80 V was maintained constant, until the sample had passed from the stacking to the resolving gel. At that moment, the voltage was increased to 150 V, and kept constant to the very end. The visualization of the samples was performed by staining the gels with the staining solution (50% (v/v) Ethanol; 9.2 % (v/v) glacial acetic acid; 0.26% (w/v) Coomassie Brilliant Blue R-250, in H_20) for 30 min at RT. The excessive background was eliminated by subsequent incubation with the destaining solution (20% (v/v) Ethanol; 10 % (v/v) glacial acetic acid, in H_2O) overnight.

3.2.2. Native gel electrophoresis

In native PAGE, the mobility depends on both the protein's charge and its hydrodynamic size, therefore being an excellent tool for detecting: (a) oligomers and aggregates; and (b) the binding events (protein complexes). The samples were mixed with 6x native loading buffer (50 mM Tris-HCl pH 7.6; 2mM

EDTA; 25% glycerol; 0.01% (w/v) bromophenol blue, in H_20), until the buffer´s final concentration of 1x, and loaded onto 6-9 % polyacrilamide gels. Electrophoresis was run using the native electrophoresis buffer (25 mM Tris pH 8.3; 250 mM glycine, in H_20) at the constant voltage of 60V. The visualization of the samples was performed by staining the gels with the staining solution for 30 min at RT. The excessive background was eliminated by subsequent incubation with the destaining solution overnight.

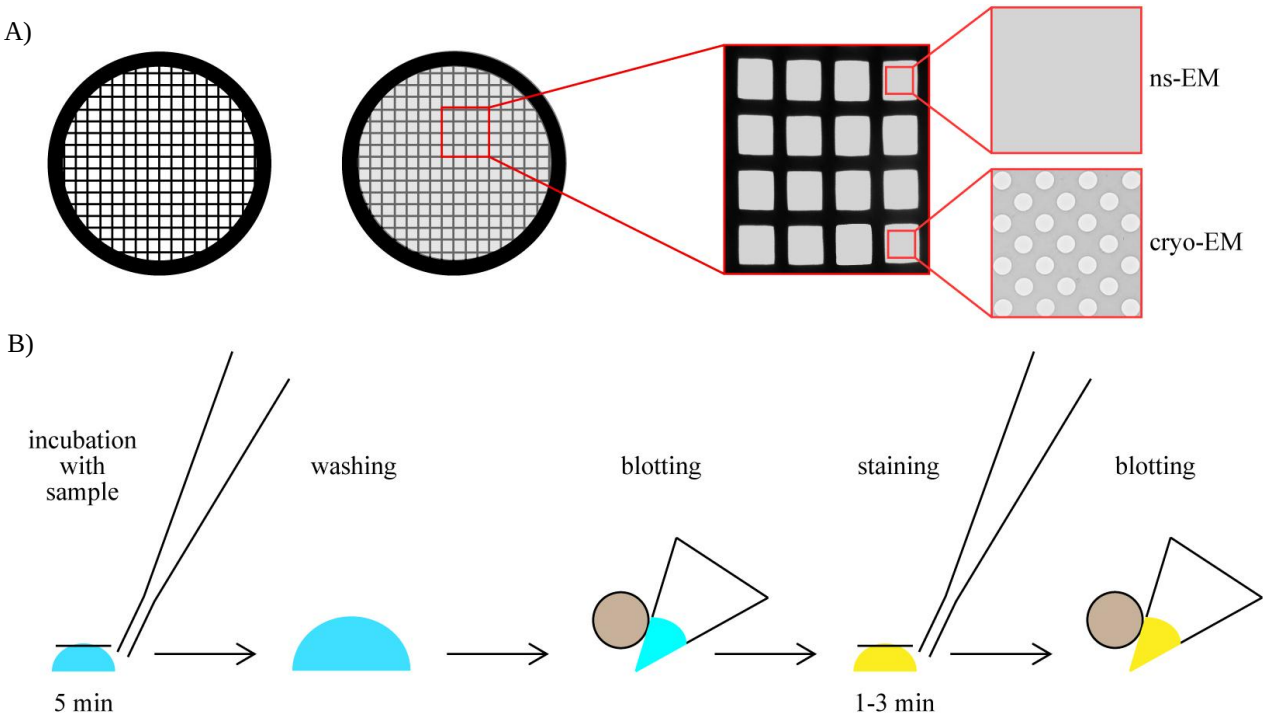
3.2.3. Biotin-streptavidin labeling

The interaction between biotin and streptavidin was employed to identify the positions of 5' ends of template and non-template strands in the studied protein:DNA complexes. The biotin labeled oligonucleotides were mixed at equimolar ratios with the corresponding, partially complementary non-labeled strands (IC–template and nontemplate IC-DNA oligomers; EC-template, non-template EC-DNA oligomers and RNA), incubated for 3 min at 92 ^{o}C, and let slowly cool down for 1h at RT. In the case of quasi-IC and EC, the annealed IC and EC DNAs were added to the sc-mtRNAP at equimolar ratio and incubated for 90 min at RT. The formed complexes were subsequently diluted in the sample buffer B mixed with equimolar amount of streptavidin and incubated for 45 min at RT. The same protocol was essentially followed for labeled-IC preparation, except that the IC-DNA was added to equimolar amounts of sc-mtRNAP and Mtf1.

3.3. Sample preparation for electron microscopy

3.3.1. Negative-staining

Copper/Rhodium (Cu/Rh) grids with 400 mesh (*MAXTAFORM* Cu/Rh HR26) were previously covered with a 6-8 nm thick carbon layer through the process known as carbon shadowing (Figure 12A). Due to the hydrophobic properties of applied carbon layer, the carbon-coated grids were subjected to an ionic discharge (30 s), few minutes before incubation with a sample. This made the applied carbon layer more hydrophilic and more prone to interaction with the sample. Grids with the carbon facing downwards were placed on the surface of 5 µl sample drop (concentration ~ 0.025-0.015 mg/ml), incubated for 5 min and washed with the adequate sample's buffer or staining agent. The staining agent, instead of buffer, was used when more homogeneous staining was required (Miyata, *et al.*, 2004). The excess of the sample and liquid was removed by blotting with *Whatman* filter paper, and the grids were subsequently stained by incubation with 5 µl drop of 2% uranyl acetate, for 1-3 min, followed by blotting with *Whatman* filter paper (Figure 12B). The whole procedure was performed at RT. Samples were allowed to dry for 30 min prior the observation.



C)

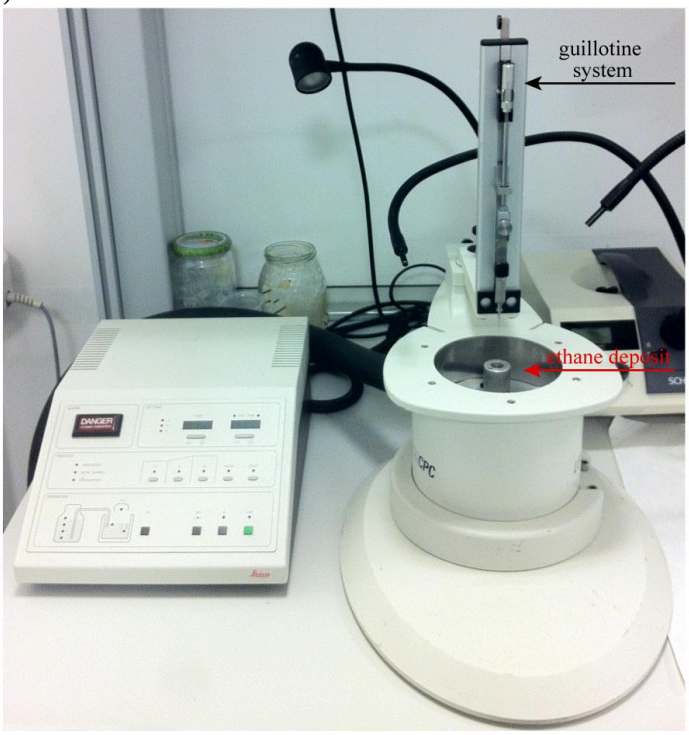


Figure 12: Sample preparation for EM. A) The Cu/Rh and/or quantifiol grids are shown before and after the carbon shadowing. The zoomed images on the right show the difference between the two types of grids, where the quantifoil grids present holes, that in this work were covered with carbon. **B)** Schematic representation of general negative-staining protocol: (1) 5 min incubation with the sample; (2) blotting the excess of the sample and the liquid; (3) 1-3 min staining with 2% uranyl acetate; (4) removal the excess of the stain by the blotting and (5) the air-drying of the grid at RT **C)** The cryopreparation chamber with guillotine (black arrow) and ethane deposit (red arrow) was used for cryo-EM sample preparation. The sample was applied to carbon-coated quantifoil grid, and the excess of the liquid was removed by blotting prior the plunge freezing in the ethane. The ethane's excess was removed before the storage in liquid nitrogen.

3.3.2. Cryo-electron microscopy

Samples for cryo-electron microscopy (cryo-EM) were applied on quantifoil grids, which have perforated support foil (*Quantifoil* R 1.2/ R1.3 300 mesh grids; reference number Q09684), previously covered with 4nm thick layer of carbon, as described before (Figure 12A). As in the case of the grids for negative-staining, these grids were subjected to a glow discharge. The grids were mounted on the plunger. Subsequent incubation with the sample for 5 minutes and blotting of the excess of the liquid/sample were followed by vitrification in liquid ethane. These steps were performed using a *Leica* CPC plunger (Figure 12C). The prepared grids were kept in liquid nitrogen until being used.

3.4. Electron microscopy

3.4.1. Image acquisition and digitalization

The samples were observed either with a *Tecnai* F20 transmission electron microscope at 200 kV (cryo-EM), or at 100 kV in a *JEOL* JEM 1230 (negative-staining). The cryoEM images were taken under low electron dose conditions (approximately 10 $e^{-}/Å^{2}$) and at ~ 110000 magnification. Defocus values were in the range of -2 to -4 µm. Images were recorded using 16 Mpixel (px) *FEI* Eagle CCD camera with 15 µm px size, thus the original sampling rate of acquired images was 1.37 Å/px. The same low electron dose conditions were followed for the acquisition of images of negatively stained samples. Negative-staining micrographs were recorded on Kodak-electron SO-163 film at 60000x magnification. Subsequent digitalization was performed using the Photoscan TD *Zeiss-Intergraph* scanner with 14µm px step size. Therefore, the original sampling rate of scanned negative film images was 2.33 Å/px.

3.4.2. Image preprocessing, particle selection and two-dimensional analysis

The phase contrast, the dominant contrast mechanism for thin stained and unstained specimens, arises from the phase shift between the scattered and unscattered electron waves (for a brief explanation of image formation see Appendix 1) (Amos *et al.*, 1982). In an ideal situation, in the presence of aberrant-free electron microscope and when the object is perfectly placed in focus, phase contrast would be equal to zero (Fermat´s principle) (Amos *et al.*, 1982). Unfortunately, the reality is drastically different. Namely, spherical (Cs) and chromatic aberrations (Cc) of the lenses, astigmatism (aberrance of perfectly symmetrical objective lens field), amplitude contrast, the image being out of focus (defocus levels), wavelength of electrons and temporal and spatial coherence of the electron beam (Zhu et al., 1997), cause a phase shift of the scattered wave. This phase shift is defined by the "contrast transfer function" or CTF (Amos *et al.*, 1982). The image recording process on film or CCD also contributes to the CTF (Koeck, 2000). Thus, it is of the highest importance to apply CTF correction before submitting the obtained micrographs to further processing. This has been done with the CTFFIND3 program (Mindell y Grigorieff, 2003). The program is also used to determine the presence and the level of astigmatism (the astigmatic angle, αast) on images of untilted

specimen. Micrographs with clear presence of astigmatism and/or drift were discarded. All the image preprocessing, particle selection and two-dimensional (2D) analysis was carried out using the XMIPP software package (Marabini *et al.*, 1996), except for previously mentioned CTF correction. Cryo-EM micrographs were downsampled by a factor of 2 or 3 (EC and IC, respectively) to reduce computational demands and to improve the signal to noise ratio. Therefore, their sampling rates were 2.74 and 4.11 Å/px. The particles were manually selected (Table 3) and extracted from micrographs. The applied box size (in pixels) was in correspondence to the size of the studied complex. The extraction step was accompanied by particle normalization, necessary for further optimization and avoidance of artifact generation (xmipp_normalize program inside XMIPP program package).

Table 3: Number of selected particles for each of the analyzed complexes.

Sample	Number of selected particles	method
Quasi IC	16721	Negative-staining
Upstream labeled quasi IC	20 437	Negative-staining
Downstream labeled quasi IC	30 898	Negative-staining
IC	15917	Negative-staining
IC	15672	Cryo-EM
Upstream labeled IC	21662	Negative-staining
Downstream labeled IC	21200	Negative-staining
EC	20873	Negative-staining
EC	38210	Cryo-EM
Downstream labeled EC	19886	Negative-staining
Upstream labeled EC	13240	Negative-staining
spRNAPIV EC	17258	Negative-staining
Upstream labeled spRNAPIV EC	23848	Negative-staining

The individual particles, boxed out from micrographs, could not immediately be used for three-dimensional (3D) reconstruction due to several reasons: (a) lack of contrast (clarity and definition); (b) limited resolution; (c) often being distorted by the incident electrons and (d) containing information not originating from the observed sample (Frank, 2006). This is tackled by particles averaging and classification (Thuman-Commike, 2001). Two types of classification, maximum likelihood 2D (ML2D) (Scheres *et al.*, 2005) and clustering 2D (CL2D) (Sorzano *et al.*, 2010), were conducted over the extracted and normalized particles. ML2D classification, as a type of multi-reference refinement, merges the alignment and the classification in one iterative process. All the particles of one data set were compared to a predefined number of references, which were derived from the original set of particles. Every particle got aligned to each of the references and classified based on the expectation-maximization approach (the highest probability to be assigned to one specific class). This way, the dependence on the initial reference images was brought to minimum (Scheres *et al.*, 2005). The particles assigned to one class were averaged and the average image

was used as a new reference for the next iteration. The process lasted until there were no more migrations between reference groups and no more changes of average, average images (Scheres *et al.*, 2005).

ML2D, however, has a tendency to misclassify the images due to dependence on the signal to noise ratio among different classes and the number of images assigned to each class (Sorzano *et al.*, 2010). Therefore, it was often complemented by another classification approach, CL2D, also provided under XMIPP program package. CL2D is a multi-reference refinement approach which is based on the correntropy, a nonlinear similarity measure between two random variables, as classification criterion. Particles, however, are not simply appointed to the classes with the highest correntropy. Their classification is rather a result of the comparison with the set of correntropies of all the images assigned to one class, and the set of correntropies of all the images not appointed to the very same class. So, they are placed into the classes that maximized the probability of being better than all the images assigned and being better than all the images not assigned. CL2D also offers two clustering criterion: (1) the classical clustering criterion, shown to be quite effective in detecting large differences between the class averages, and (2) the robust clustering criterion which was specially designed for subtle differences embedded in very noisy images (images obtained by cryo-EM) (Sorzano *et al.*, 2010).

Combined use of ML2D and CL2D classifications was not only employed to create references that were later on used for generation of reference models (initial models), but also to deal with heterogeneous populations of particles, such as biotin-streptavidin labeled complexes, and to discard the particles that would affect later steps of 3D reconstruction, such as the particles that were positioned too close to each other and/or particles corresponding to aggregates rather than complexes. In the case of some heterogeneous particles populations, the KerDenSom (*Kernel Density Estimator Self-Organizing Map*) program was additionally employed (Pascual-Montano *et al.*, 2001). The program is based on an algorithm for particle alignment and their separation upon the difference at a previously determined region, which is defined by the user (Frank *et al.*, 1992, Penczek *et al.*, 1992).

3.4.3. 3D reconstruction

The majority of the 3D reconstruction steps, including the generation of initial models, were performed using the EMAN program package (Ludtke *et al.*, 1999). All the present refinement program packages exhibit, to some extent, the dependence on the initial reference (model bias). This can be overcome by running several independent reconstructions starting from various initial models. There are different methodologies implemented in EMAN that can be employed to generate initial reference maps (Figure 13):

- **Common lines:** Initial model is created upon the class averages created in the previous step. This methodology is based on central section theorem (Crowther *et al.*, 1970), postulating that Fourier transformation of 2D images, representing the projections of the observed object, constitute the planes of that object in 3D Fourier space. Thus, existence of common lines between those planes can be assumed. EMAN´s *StartAny* program searches for the common lines of provided class

averages, and interpolates the rest of information upon them to generate a reference volume, used for subsequent iterative angular refinement.

- **Blob:** *Makeinital* program allows the generation of a geometric object which dimensions should correspond to averaged class images.
- **Noise:** This method is based on the previously described one. The noise, via *Proc3d* program, can be assigned to the initial model created using the *Makeinitial* program.
- In the case of streptavidin labeled complexes and complexes analyzed by cryo-EM, the previously obtained EM maps of non-labeled or negatively-stained complexes were filtered to 70 Å, and used as the initial models.

Rough experimental data (particles) cannot be directly converted into an optimal 3D model. Therefore, preliminary initial model is iteratively refined against the experimental data set in a process known as iterative angular refinement. The process starts when the initial model is projected in 3D space, where the number of projections depends on the angles value, defined by the user. The projections are used as initial references (theoretical references) against which the experimental data is aligned and classified. Then, classified particles are averaged, creating the experimental references. As the experimental references carry the information about their spatial orientation, from the theoretical references from which they are derived, they can be simply inserted in Fourier space to create a new model. The new model is used as the initial model for the next iteration. The whole process is repeated until a satisfactory stable model is obtained, and when significant changes cannot be observed between iterations (Figure 13) (Ludtke *et al.*, 1999).

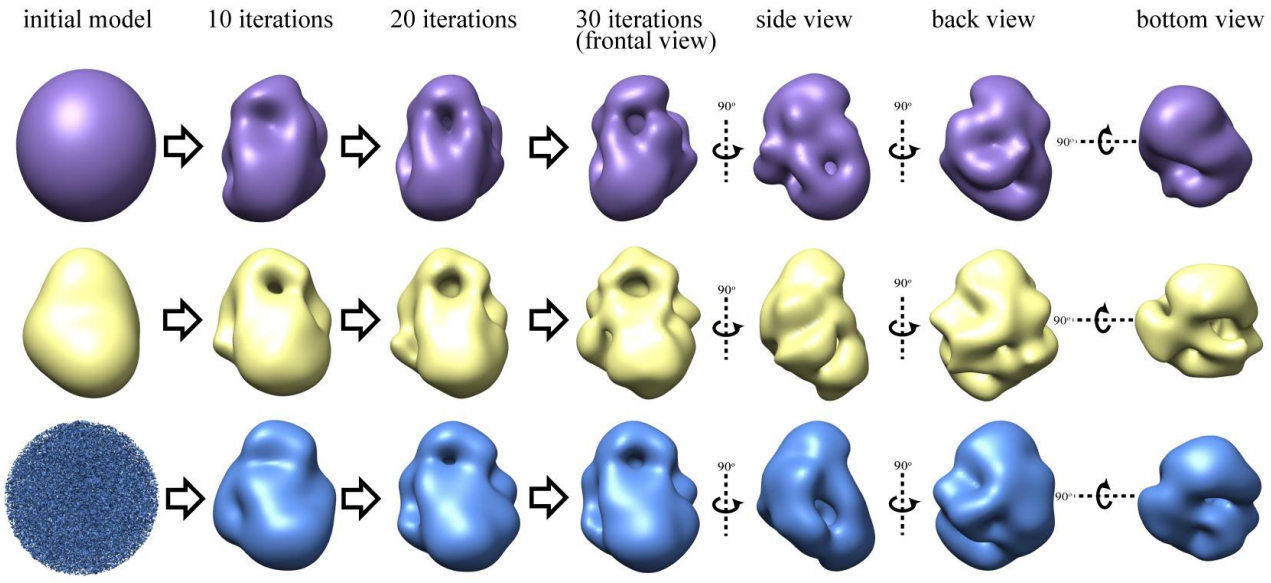


Figure 13: The representation of iterative angular refinement. The different processing steps in EMAN starting from 3 different initial models: blob (purple), common lines (yellow) and noise (blue) are shown for IC. The necessity for simultaneous independent reconstructions is the consequence of a model bias that is present in all existing program packages. Thus, they represent a kind of validation mechanism for the model that will be submitted to the final steps of processing in XMIPP.

Once the independent reconstructions, starting from different initial models, converged to a similar EM map, the final steps of 3D reconstruction were approached, carried out in XMIPP software package. The

stable model from EMAN was filtered to 60 Å and used as the initial model for Projection Matching. However, in the case of heterogeneity (EC cryo-EM; streptavidin-biotin labeled ICs) the most stable model was submitted to maximum likelihood 3D classification (ML3D) to address this problem (Scheres, 2010). The separated particles were subjected to processing in EMAN and the whole 3D reconstruction cycle was repeated *de novo*.

The programs implemented in EMAN and XMIPP were used to estimate the resolution of the final models. Namely, the whole set of particles forming the final EM map was randomly divided in two groups and used to build two independent models. The generated models were compared in Fourier space and calculated correlation was expressed in function of spatial frequency or Fourier Shell Correlation (FSC) (van Heel, 1984). The resolution was determined as the reciprocal value of FSC at which the correlation coefficient had the value of 0.5 (the 0.5 criterion). The final validation of obtained models was performed by comparison of final 3D map projections and generated class averages (EMAN) with the results of 2D classifications (Figure 14).

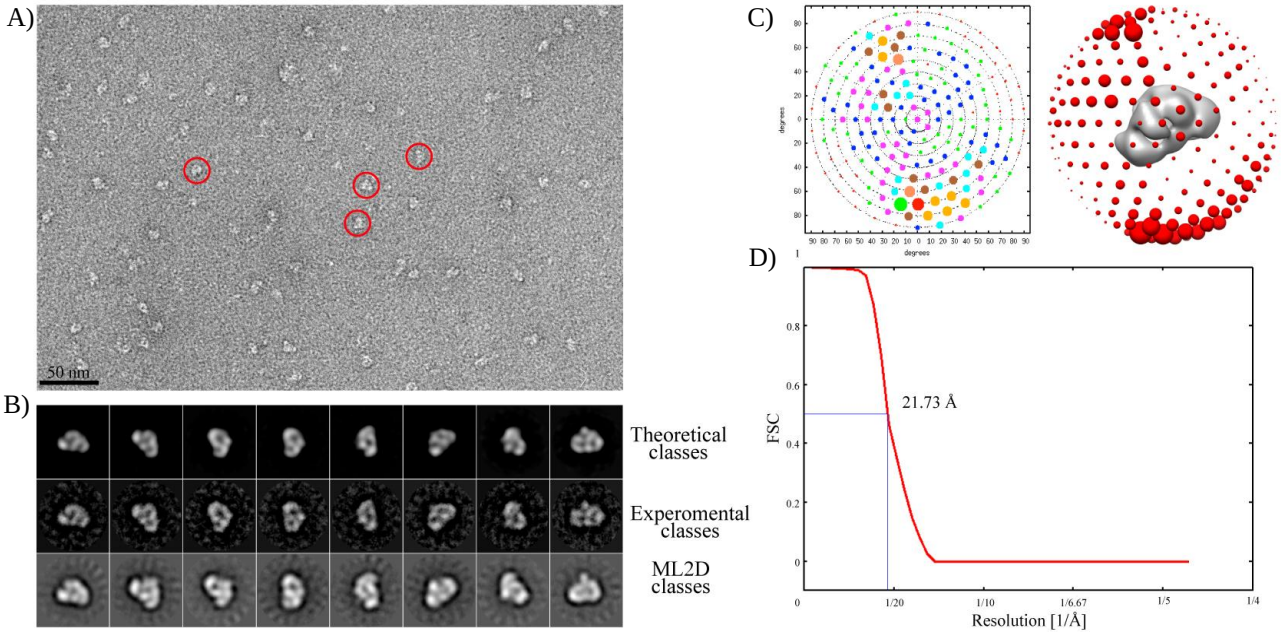


Figure 14: The additional assessment of reconstructions' quality. A) An example of a typical negative staining micrograph of the IC complex, with some selected individual particles (red circles) **B)** The quality of 3D reconstruction is represented at the example of IC, where further validation can be obtained by comparison of final 3D map projections and generated class averages (EMAN) with the results of ML2D classification. **C)** 2D and 3D representation of angular distribution of the particles that were used in the final step of 3D reconstruction of IC, performed by Projection Matching. Each circle (2D) or sphere (3D) represents one projection of the final model with defined angular values. Their size is proportional to the number of assigned individual particles. D) The computed resolution curve of IC. The resolution of the obtained model has been estimated to be 21.73 Å at the 0.5 criterion.

3.4.4. Segmentation and fitting of atomic models and structures

The comparison of the final models and the fittings of human mtRNAP´s atomic structure were preceded by segmentation of adequate 3DEM density maps in USCF *Chimera* (Pettersen *et al.*, 2004). The fittings were firstly preformed manually and then rendered by *Chimera*, using a real time correlation criterion between the reference model (final model) and the surface created from the atomic structure/model.

4. Results

4.1. Structural characterization of quasi-IC

4.1.1. 3D reconstruction of negatively-stained quasi-IC

The first state of the sc-mtRNAP that we sought to characterize was the quasi-IC. In this state, the polymerase is bound to the bubble promoter but lacks the Mtf1 factor. The sc-mtRNAP is able to start the transcription from the negatively supercoiled and bubble promoters which energetic barrier of DNA opening has been significantly lowered (Matsunaga y Jaehning, 2004, Tang *et al.*, 2009a). Prior to the structural characterization, the purity of the complex was assessed by SDS-PAGE, which shown the absence of any notable contamination and degradation of quasi-IC. A slight smear at the native-PAGE indicated the presence of not only the monomeric form of the complex, but also the oligomeric species and even aggregates (Figure 15).

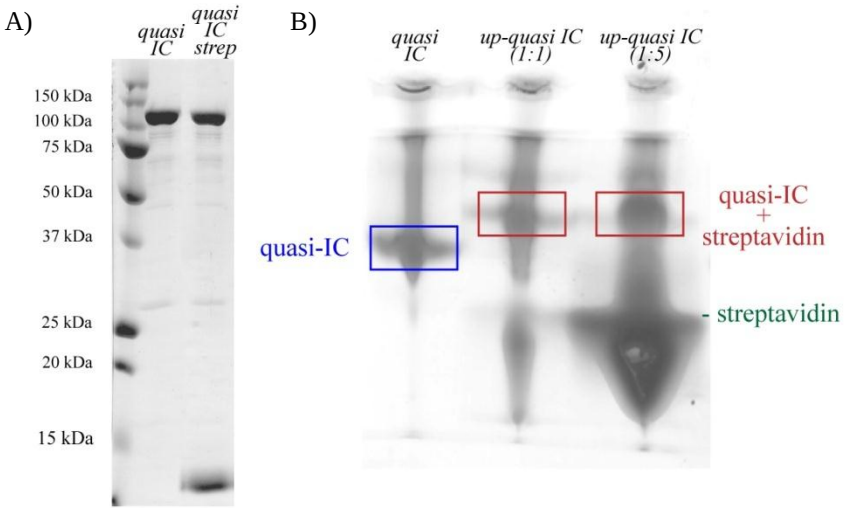


Figure 15: Electrophoretic analysis of quasi-IC and upstream labeled quasi-IC. A) 10% SDS analysis of quasi-IC and upstream streptavidin labeled quasi-IC used for EM studies. Sc-mtRNAP has a molecular mass of 153 kDa and the absence of any additional signal is indicator of the high purity of the analyzed sample. **B)** 9% Native gel: Lane named the quasi-IC corresponds to quasi-IC, while the lanes marked as up-quasi IC (1:1) and up-quasi IC (1:5) correspond to upstream labeled complexes at 1:1 and 1:5 molar ratios. The presence of the signals at the stacking gel reveal the existence of aggregates, while the smear at quasi-IC lane gives information of the formation of differently sized oligomers of IC. The addition of streptavidin in a higher amount doesn't significantly increase the level of formed labeled quasi-IC, but it drastically contributes to the heterogeneity of the analyzed sample.

This implied that more care should be taken at the moment of particle selection and at the same time the employment of more extensive 2D classification. Images exhibiting the presence of more than one particle, and/or images that were extremely bright or dark, were removed. This resulted in a decrease in the number of originally 16721 selected particles to 12848. The classes that were common for different ML2D and CL2D classifications were used to obtain general the dimensions of the complex, necessary to subsequently create a blob and a noise, and to generate the initial model employing the common lines approach. Once the independent reconstructions in EMAN gave similar results, the more stable model was filtered to 60 Å and subjected to the processing in XMIPP (Projection Matching). The calculated resolution of the final 3D model originating from Projection Matching is around 19 Å using 0.5 FSC criterion.

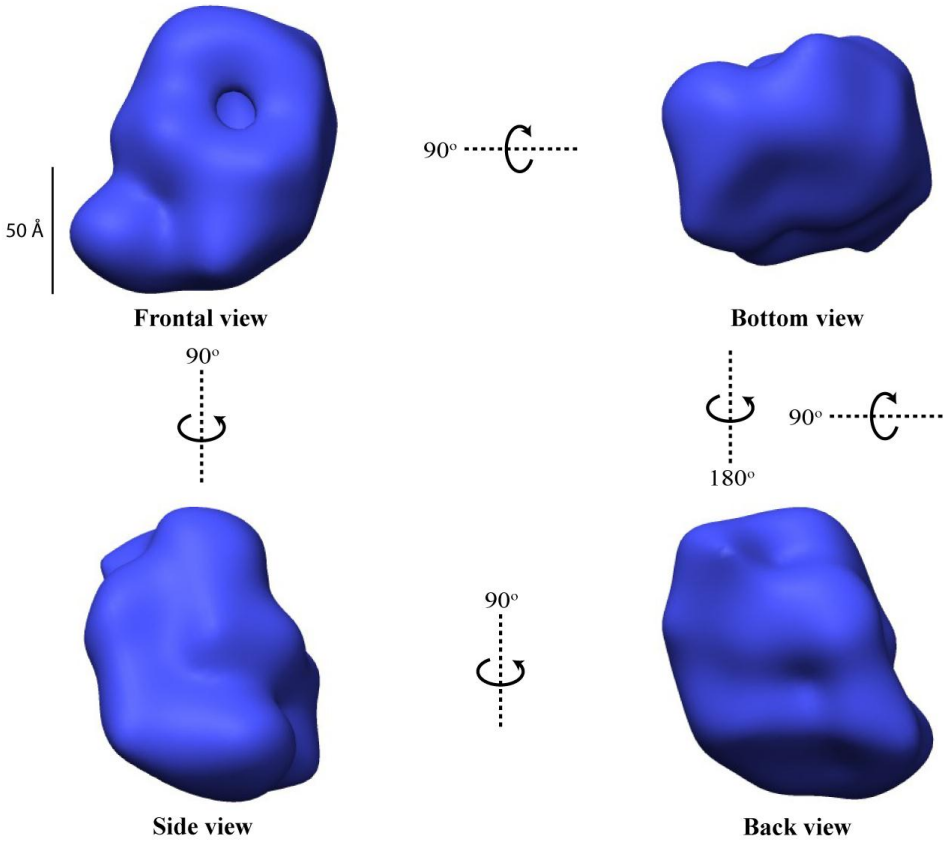


Figure 16: 3D reconstruction of quasi-IC. The final EM map of quasi-IC at 19 Å resolution presents an overall shape of a cupped right hand, of 102 Å in height, 74 Å in width and 63 Å in depth. The model is presented in four characteristic views: frontal, side, back and bottom. The scale bar represents the length of 50 Å.

The 3D reconstruction of quasi-IC revealed a compact structure with 102 Å in height, 74 Å in width and 63 Å in depth. Even though the low resolution of the 3D reconstruction does not permit to clearly distinguish the different subdomains, the general shape of a right cupped hand, characteristic for the members of single-subunit RNAP family, can be observed. As described in header 1.3.2.1 it is thought that all the members of this family share the similar structure with the T7 RNAP. The structure of this enzyme can be divided into two parts: the NTD which role is in the promoter recognition, binding and melting; and the core, which is composed of the thumb, the palm, and the fingers subdomains. The mtRNAPs share 8 regions of homology with T7RNAP that are mostly distributed in the core. The structural characterization of the human mtRNAP (Ringel *et al.*, 2011) has confirmed the structural similarity of the mtRNAPs and T7RNAP. The mtRNAPs' subdomains have been named according to the exhibited homology with the T7RNAP: the core (residues 648-1230) with the fingers, the palm and the thumb subdomains, the NTD (residues 369- 647) which contains the AT rich recognition loop and the intercalating β hairpin; and the N-terminal extension which can not be found in the phage homologue. The opening (Figure 16) is assumed to be a possible exit channel for the RNA. The X-ray atomic structure of human mtRNAP (PDB entry 3SPA) was fitted into the EM map to obtain more information about the architecture of quasi-IC (Figure 17). Firstly, the docking was performed manually and then refined using Chimera. The fitting gave further support to our reconstruction as all the elements of the atomic structure were fitted fairly well in the EM map. The small portions of alpha helixes of the palm, the fingers and the N-terminal domain seen to be localized outside the

EM map, are connected to the rest of the atomic structure by loops indicating their flexibility and therefore they might assume different position in the quasi IC complex (Figure 17).

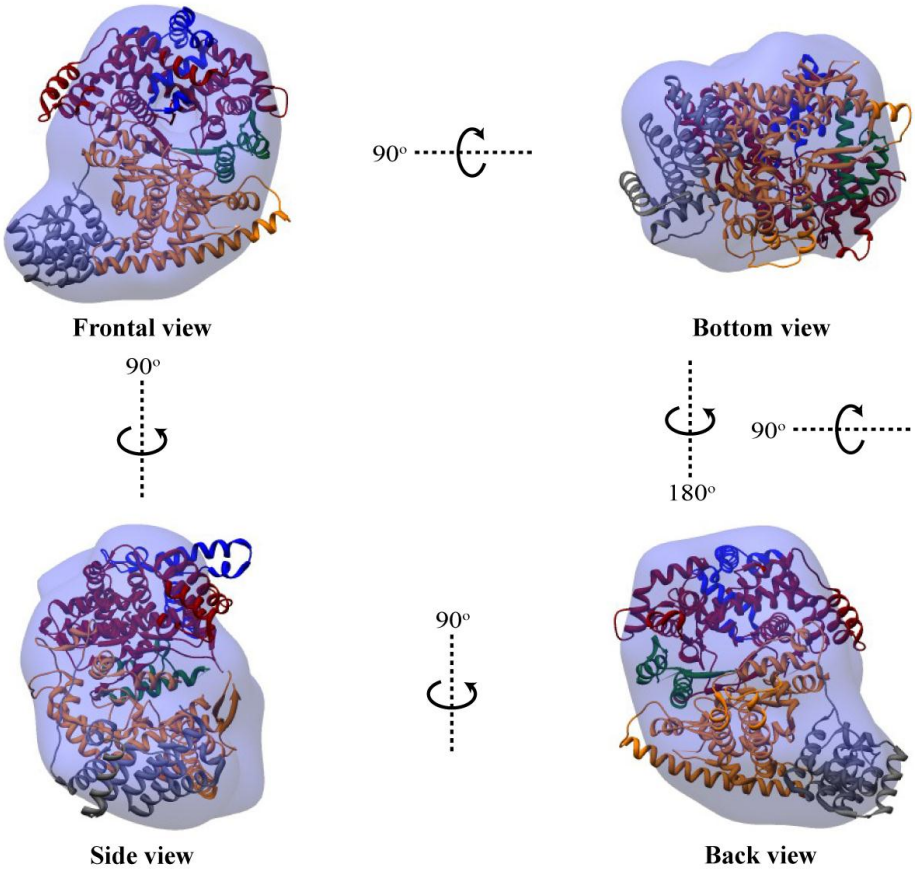


Figure 17: Docking of human mtRNAP's X-ray structure into quasi-IC 3DEM map. Regions exhibiting sequence homology with T7 RNAP are emphasized: NTD (orange); Thumb domain (dark green); Palm domain (dark red); Fingers (blue); mtRNAPs contain an additional N-terminal extension (grey) which is composed of MTG; region involved in transcription-translation coupling and putative pentatricopeptide repeat (PPR), found in proteins implicated in RNA-processing in mitochondria and chloroplasts. The docking is shown in four typical views: frontal, side, back and bottom.

4.1.2. Localization of upstream end of DNA in quasi IC-complex

The upstream streptavidin labeled quasi-IC was formed as previously described (Header 3.2.3). The addition of an excess of streptavidin does not seem to induce any significant change in quasi IC migration, suggesting that the quasi-ICs are effectively saturated at these conditions (Figure 15). However, the presence of a large excess of free non-bound streptavidin would definitely contribute to the formation of aggregates due to their multivalent nature, making the EM studies of such sample more complicated. For this reason, the samples used for the EM analysis were those incubated with streptavidin at 1:1 ratio (Figure 17, lane 2). Combination of the presence of differently sized oligomeric species of sc-mtRNAP, of streptavidin oligomers, of the small but notable portion of non-labeled sc-mtRNAP, and similar dimensions of one of the projections of quasi-IC and streptavidin, yielded for stricter particle picking and in-depth 2D classification, employing also KerDenSom in an attempt to deal with the evident heterogeneity at the 2D level. This was complemented by the ML3D classification in order to separate at 3D level the particles that corresponded to non-labeled quasi IC from the ones of the labeled complex, as the previously obtained EM map of quasi-IC

was filtered to 60 Å and used as the initial model. This reduced the originally collected particle data set of 20437 particles to 16614 that were subsequently used for 3D reconstruction in EMAN and XMIPP. Also, the EM map of quasi-IC was used as an additional initial model for the iterative angular refinement in EMAN. Once the four independent 3D reconstructions, starting from the blob, noise, common lines and previously mentioned filtered quasi-IC EM map converged to a similar EM map, the process was carried out in XMIPP. The estimated resolution using 0.5 FSC criterion is similar to the one for non-labeled quasi IC (~ 18 Å). The streptavidin is a tetrameric protein of 53 kDa, comprising 1/4 of the mass of the formed complex, thus expected to be easily observed as an extra mass. As a matter of fact, the 3DEM map of labeled quasi-IC shows the existence of two distinguishable electro densities: the larger one, which shape and the presence of the entrance to the centrally positioned cavity can be easily identified as the quasi-IC, and a smaller one, which shape and size corresponds to the streptavidin tetramer (Figure 18A). However, the lack of resolution due to the limitations of the method prevents the more precise localization of the upstream DNA. The streptavidin binding to the labeled upstream DNA might cause a slight dislocation of a flexible NTD in quasi-IC, exhibited through the small portion of N-terminal extension placed outside of the EM map (Figure 18B).

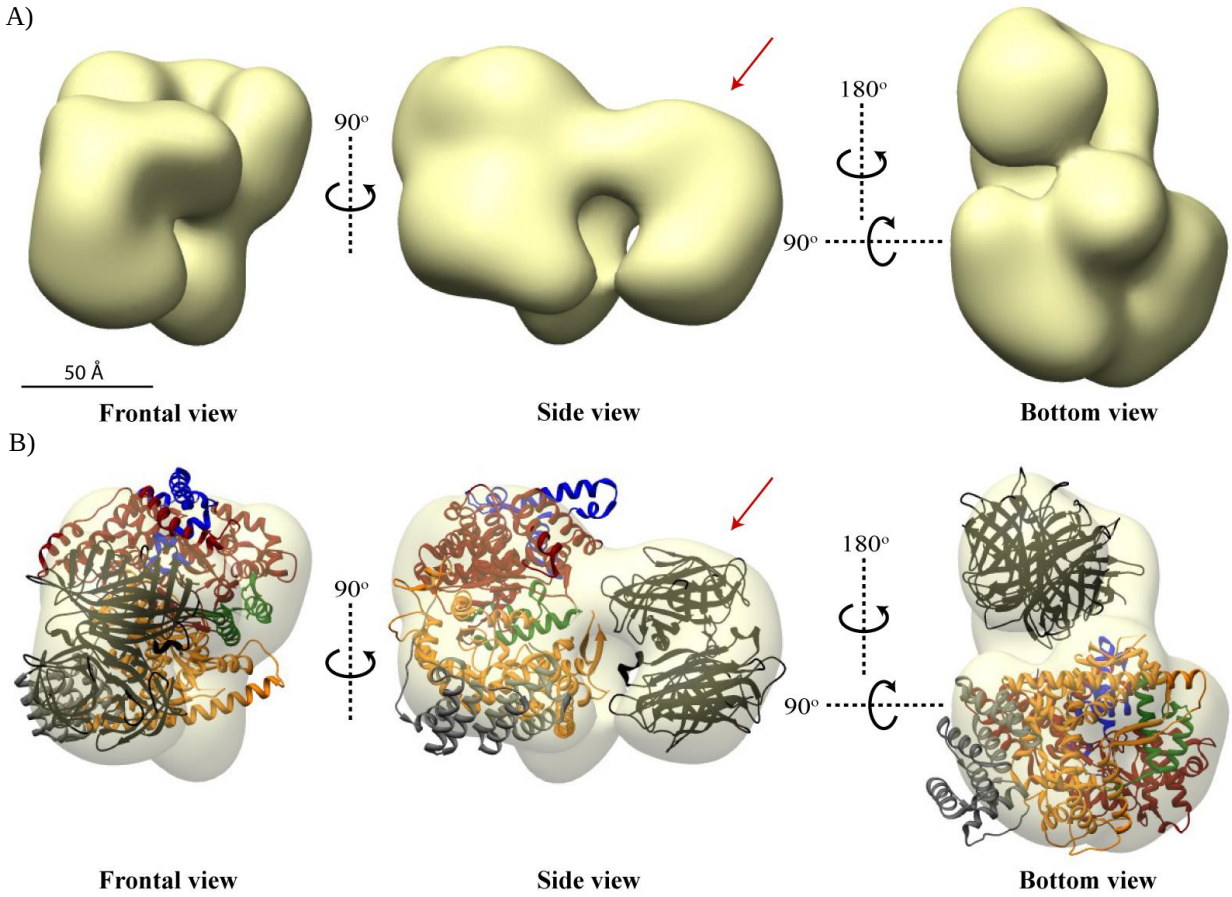


Figure 18: 3D reconstruction of upstream labeled quasi-IC. A) The final map originating from Projection Matching is shown in three characteristic views: frontal, side and bottom. The red arrows indicate an extra mass connected to the frontal part of the sc-mtRNAP that corresponds to the bound streptavidin, indirectly identifying the position of the upstream end of the IC-DNA. The scale bar represents the length of 50 Å. **B)** Docking of human mtRNAP X-ray structure into quasi-IC EM map. The subdomains of human mtRNAP are shown: NTD (orange); Thumb domain (dark green); Palm domain (dark red); Fingers (blue); and the additional N-terminal extension (grey). The atomic structure of streptavidin (PDB entry 1N43) is dyed in black. The docking is shown in the same views as A.

4.2. Structural characterization of IC

4.2.1. 3D reconstruction of negatively-stained IC

As previously shown, sc-mtRNAP and Mtf1 interact at 1:1 molar ratio to form a ~200 kDa functional unit capable of promoters' DNA melting and initiation of transcription (Tang *et al.*, 2009a). The sc-mtRNAP alone can bind to the promoter and non-promoter DNA, but those complexes exhibit high instability and are rapidly disassembled (Deshpande y Patel, 2012, Kim *et al.*, 2012). In order to be able to compare the quasi-IC and IC we have used the bubble promoter to also form the latter. This means that we have been dealing with IC at the same open promoter state (RPo). The native gels (Figure 19) have shown the increased stability of IC comparing to the quasi-IC, which can be assigned to the combined stabilization effect of Mtf1 alone and possible conformational changes induced in mtRNAP that exposed additional surfaces, which might be involved in the binding of DNA. The increased stability has also been observed under EM as one predominant homogeneous population of particles.

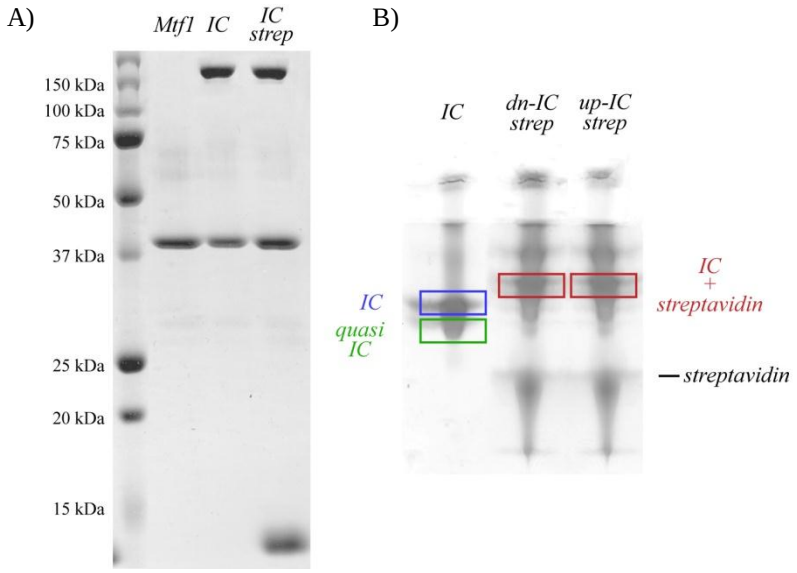


Figure 19: Electrophoretic analysis of IC and streptavidin labeled IC. A) 12% SDS gel: 1st unlabeled lane corresponds to molecular weight markers, while Mtf1, IC and IC strep to purified Mtf1, IC and streptavidin labeled IC, respectively. **B)** 9% Native gel: IC lane corresponds to IC, while lane dn-IC strep and up-IC strep to upstream and downstream labeled complexes at 1:1 molar ratio. The IC lane shows the presence of two signals that correspond to IC and quasi-IC (framed in blue and green), while the upstream and downstream labeled ICs (framed in red) exhibit a shift in migration due to the presence of streptavidin. The non-bound streptavidin is seen as a smear at the lower part of the gel.

Samples were subjected to a similar EM analysis, which allowed the obtaining of a 3D map at 21.73 Å of resolution. The comparison of this map with the one previously obtained for quasi-IC reveals the presence of a round protruding mass located at the bottom of the frontal view, which can be assigned to Mtf1 (Figure 20, red arrow).

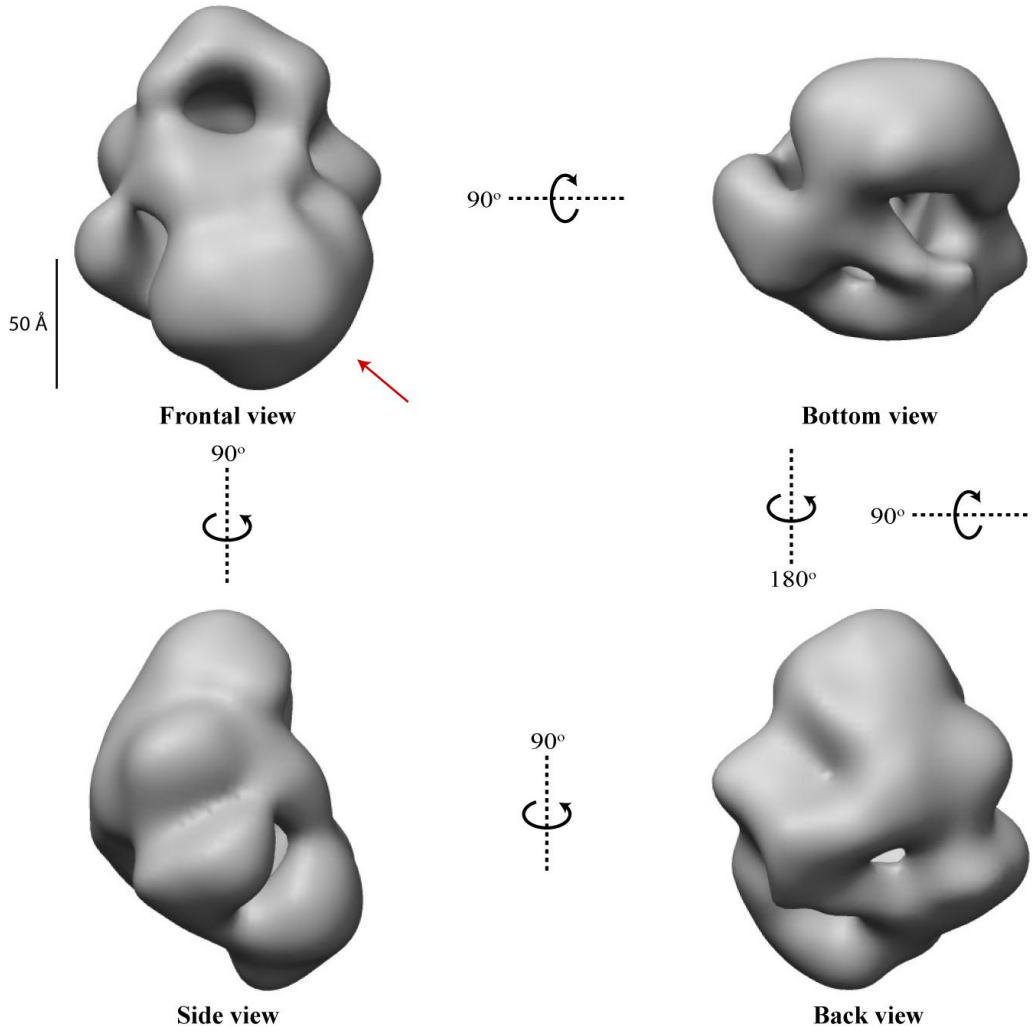


Figure 20: 3D reconstruction of IC. The final map of sc-mtRNAP:mtf1:bubble IC-DNA and therefore of the Rpo (open promoter IC) is shown in the same characteristic views. The red arrow at the front view indicates towards the mass formed by Mtf1. The final resolution of the model is ~ 22 Å. The scale bar represents the length of 50 Å.

To allow a more precise identification of the two components, IC 3DEM map was segmented using UCSF Chimera, resulting in its division into the sc-mtRNAP and Mtf1 portions, later used for the human mtRNAP and Mtf1 atomic structures fitting. The docking shows for the first time the spatial organization of IC, placing the Mtf1 at close proximity of N-terminal domain, the palm and the fingers subdomains of sc-mtRNAP (Figure 21). The positively charged central cleft of Mtf1, placed between the N-terminal α/β domain and C- terminal helical domain l, has been indicated, by the docking, to be facing towards sc-mtRNAP. The 3DEM model of IC, is wider than the one of quasi-IC due to the presence of an additional mass (Figure 21, mass indicated by the black arrows) that might be formed by the upstream DNA and partially by the C-terminal tail of Mtf1, which could be organized in the presence of DNA. N-terminal extension of human mtRNAP atomic structure (first 300 amino acids), which contains the MTG and sequences shown to be involved in transcription-translation coupling, in autoregulation and in establishing interactions with Mtf1 (Rodeheffer *et al.*, 2001, Bryan *et al.*, 2002, Rodeheffer y Shadel, 2003, Markov *et al.*, 2009), fills perfectly the small region placed at the entrance of the central cavity and at the close proximity of the N-terminal α/β domain of Mtf1 (Figure 21, frontal view; mass indicated by the red arrow).

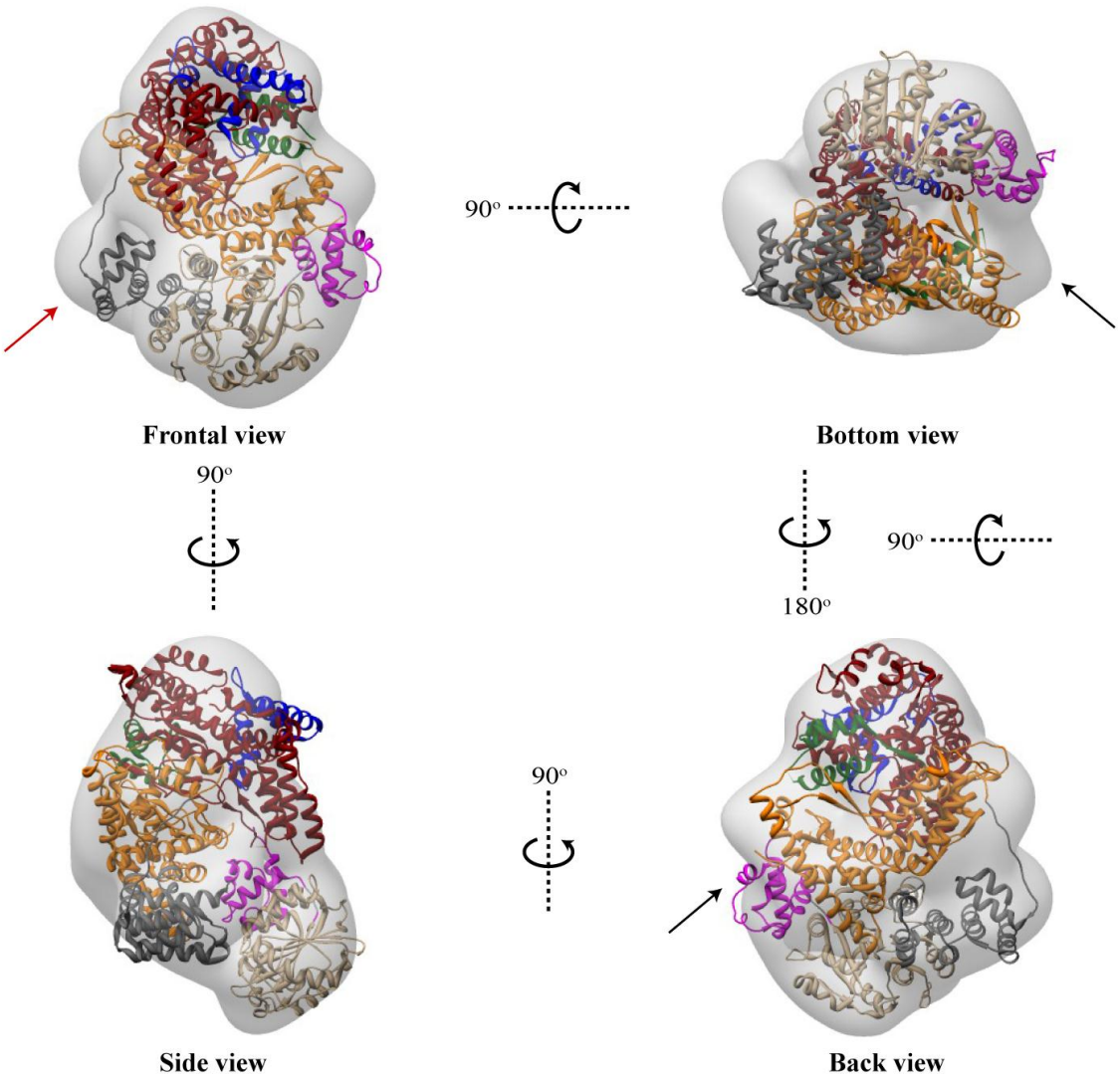


Figure 21: Docking of the atomic structures of human mtRNAP and Mtf1 into 3DEM map of negatively-stained IC. The subdomains of human mtRNAP are shown: NTD (orange); Thumb domain (dark green); Palm domain (dark red); Fingers (blue); and the additional N-terminal extension (grey). The N-terminal α/β domain and C- terminal helical domain l of Mtf1 are dyed in tan and magenta, respectively. The black arrows points out the extra mass that can't be seen in the 3DEM map of negatively-stained quasi-IC and that might be formed by C-terminal tail of Mtf1 and the IC-DNA. The red arrow indicates to the small region in which the N-terminal extension of the atomic structure can be perfectly fitted.

4.2.2. Cryo-EM of IC

IC complex is a good candidate for its cryoEM analysis due to its high stability and evident homogeneity, together with a molecular mass of 210 kDa (~ 200 kDa of protein component + ~ 10 kDa contributed by IC-DNA). Quantifoil grids with IC sample were prepared, as explained before (see Header 3.3.2), and observed under cryo conditions.15672 particles were manually selected and classified using the CL2D and ML2D. The particles assigned to the classes with a very low signal to noise ratio were removed for the further processing. Remaining 10893 particles were used to generate the cryo-EM map of IC, at 15 Å resolution, that confirmed our previously obtained negatively-stained model (Figure 20). Two main portions can be distinguished keeping the same relative positions to each other as in the case of negatively-stained 3DEM map of IC (Figure 22).

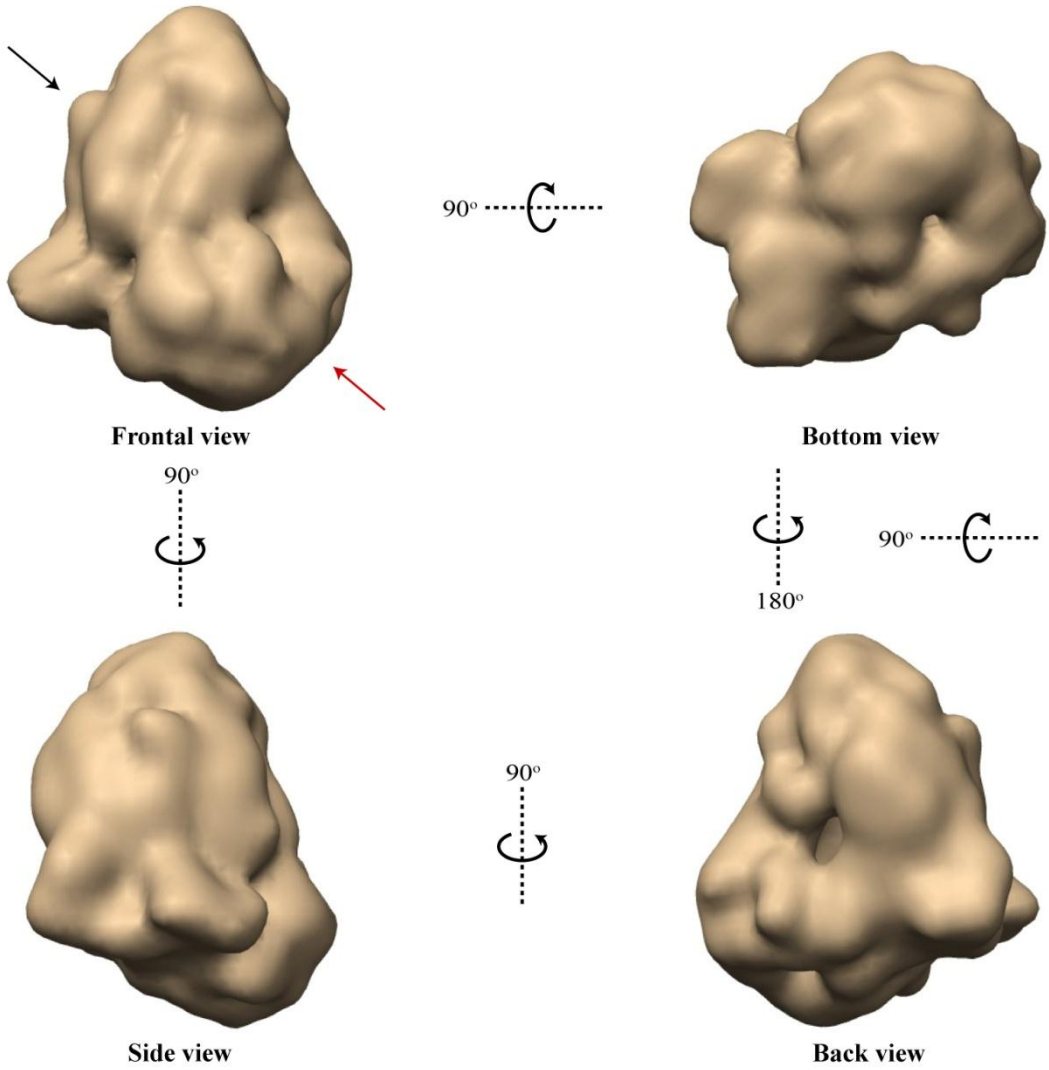


Figure 22: 3D reconstruction of IC obtained by cryo-3DEM. The same views are shown as in the case of other 3DEM maps: frontal, side, back and bottom views. Two main masses can be observed, the larger one corresponding to mtRNAP (black arrow) and smaller one to Mtf1(red arrow).

As the cryo-EM deals with protein samples maintained at their physiological or close to their physiological conditions, the generated cryo-EM models are more reliable than the ones obtained through the negative-staining. This is often complemented with the higher resolution, allowing a more precise fitting of available atomic structures. In our particular case, the N-terminal extension (Figure 23, portion of the atomic structure dyed in grey) and the frontally positioned alpha-helical bundle of the palm subdomain (Figure 23, portion of the atomic structure dyed in red) are in the close proximity of N-terminal α/β domain and positively charged cleft of Mtf1, respectively, as it was also seen in the case of 3DEM of negatively-stained IC (Figure 21). The NTD (Figure 23, portion of the atomic structure dyed in orange) seems to be placed near the C- terminal helical domain l and a small mass, which is not occupied by the atomic structure (Figure 23, back and bottom views, indicated by the black arrows). This mass is thought to be formed by the C-terminal tail of Mtf1 which is absent in the atomic structure, and the upstream part of the DNA.

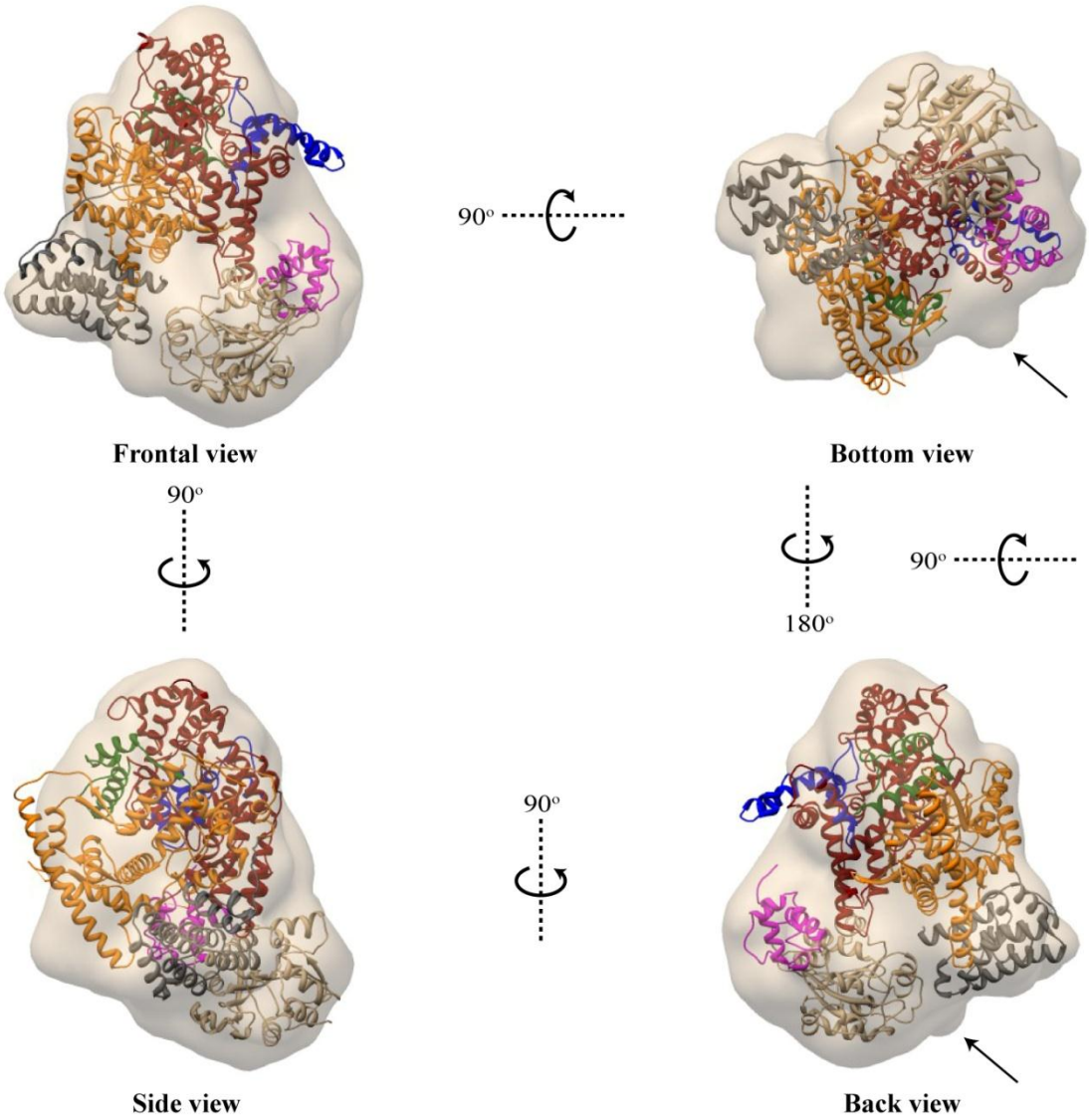


Figure 23: Docking of atomic structures of human mtRNAP and Mtf1 into cryo-3DEM map of IC. The subdomains of human mtRNAP are shown: NTD (orange); Thumb domain (dark green); Palm domain (dark red); Fingers (blue); and the additional N-terminal extension (grey). The N-terminal α/β domain and C- terminal helical domain l of Mtf1 are dyed in tan and magenta, respectively. The black arrows indicate the mass that might be contributed by the C-terminal tail of Mtf1.

4.2.3. Localization of upstream and downstream ends of DNA in IC

The next aim in our research was to give an answer to the question about the origin of the extra region visible at the back of 3DEM map of negatively-stained IC (Figures 20 and 21), and to the question if there is a difference in the position of the DNA when sc-mtRNAP is present alone or in complex with Mtf1. In the case of T7 RNAP polymerase, the sequence of interactions between the AT rich recognition loop, the specificity loop and the intercalating β-hairpin with the adequate regions of Class III T7 promoters, induces the bending of DNA for 40-60^{o} around the nt at -1 position (Deshpande y Patel, 2012). The requirement for Mtf1 implies that sc-mtRNAP, despite being capable of binding to the promoter sequence alone, should have distinct behavior that might be manifested through the different positioning of the promoter DNA. The presence of evident heterogeneity (Figure 19) and the faced problems during the reconstruction of the labeled quasi-IC, led to the collection of larger particle data sets for the upstream and downstream labeled ICs,

21662 and 21200 particles, respectively. Furthermore, the similar molecular mass of Mtf1 and streptavidin affected the 3D reconstructions by generating problems regarding the determination and assignment of the correct angles to some of the projections which resulted in repetitive cycles of processing in EMAN, ML3D and Projection Matching. The 3D reconstruction of upstream labeled IC, confirmed our assumption regarding that the extra region, visible at the back of the 3DEM model of negatively-stained IC corresponds to the IC DNA (Figures 20 and 21). The mass corresponding to the streptavidin protrudes from this region (Figure 24A and 24B). Above all, they are connected by a thin bridge structure that is formed by the very upstream end of DNA. The reason for the N-terminal α/β domain of Mtf1 not being disconnected from what seems to be N-terminal extension of mtRNAP, can be assigned to a defect of negative-staining approach and different penetration of uranyl acetate in two cases.

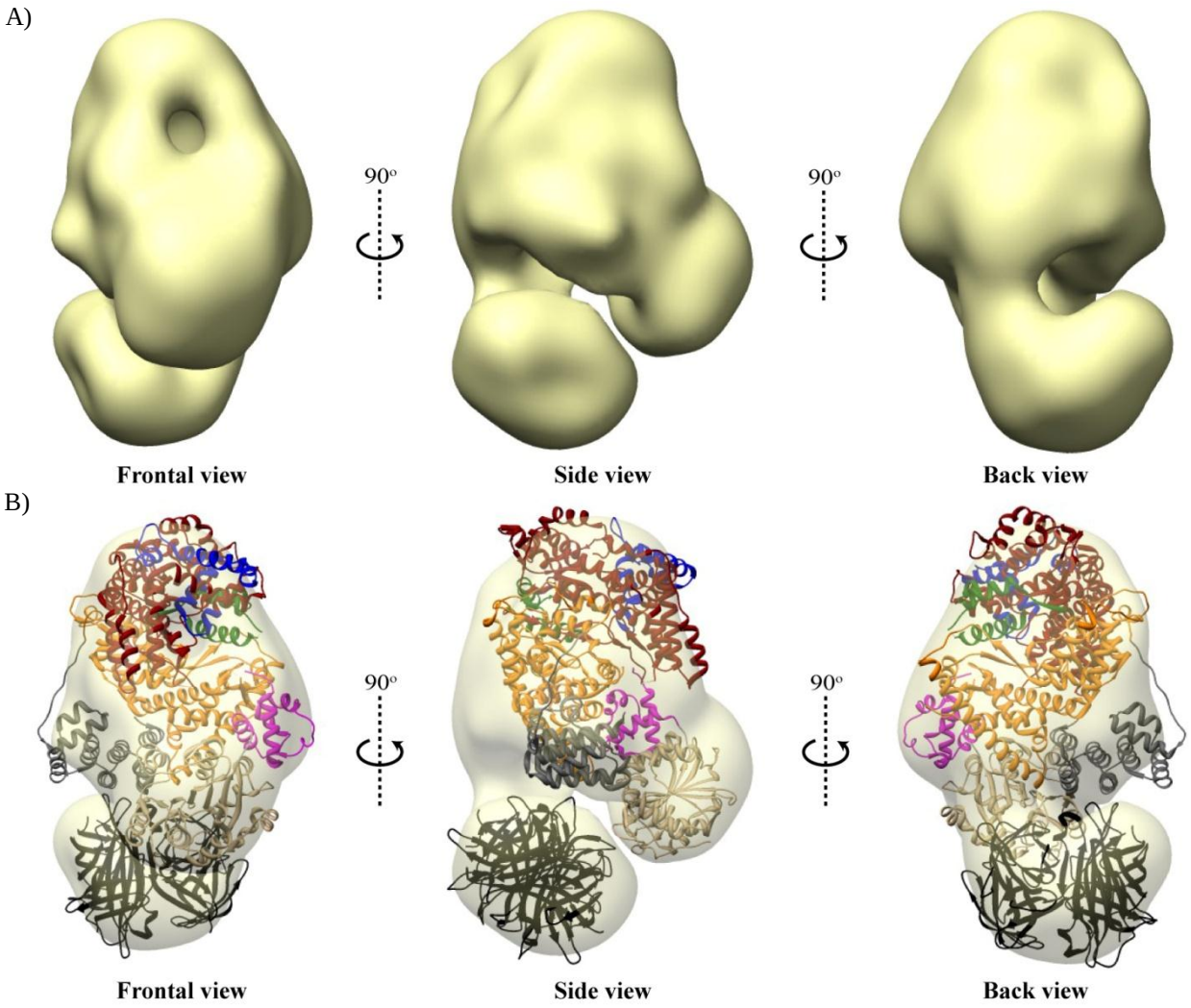


Figure 24: 3D reconstruction of upstream labeled IC. **A)** The final EM map of upstream labeled IC shown in three views: frontal, side and back. **B)** Fittings of atomic structures of human mtRNAP, Mtf1 and streptavidin. NTD, Thumb, Palm and Fingers domains and the additional N-terminal extension are colored in orange, green, red, blue and grey, respectively. The N-terminal α/β domain and C- terminal helical domain l of Mtf1, and streptavidin are dyed in tan, magenta and black.

The transition from IC to EC involves the disruption of the contacts between the T7 RNAP and promoter, while the established interactions with the downstream end of DNA remain intact. If this is also the case for mtRNAPS, we could expect the streptavidins of the downstream labeled IC and EC to be at

similar, if not the same, positions. Thus, the corresponding 3DEM maps should provide the information whether these two types of related RNAPs share a common mechanism of activity. In the 3DEM density map of downstream labeled IC, the extra density that can be attributed to streptavidin is located on the left of the sc-mtRNAP, in the area between the NTD, the N-terminal extension and the part of the palm domain. The N-terminal α/β domain of Mtf1 is positioned relatively far from this region. This implies that downstream end of DNA exclusively establishes the contacts with sc-mtRNAP. Furthermore, the tight positioning of streptavidin and the sc-mtRNAP indicates that the downstream part of the DNA is buried deeper than the upstream end in the polymerases binding cleft. This would allow the formation of more extensive interactions with DNA and stabilization of the complex (Figure 25A and 25B).

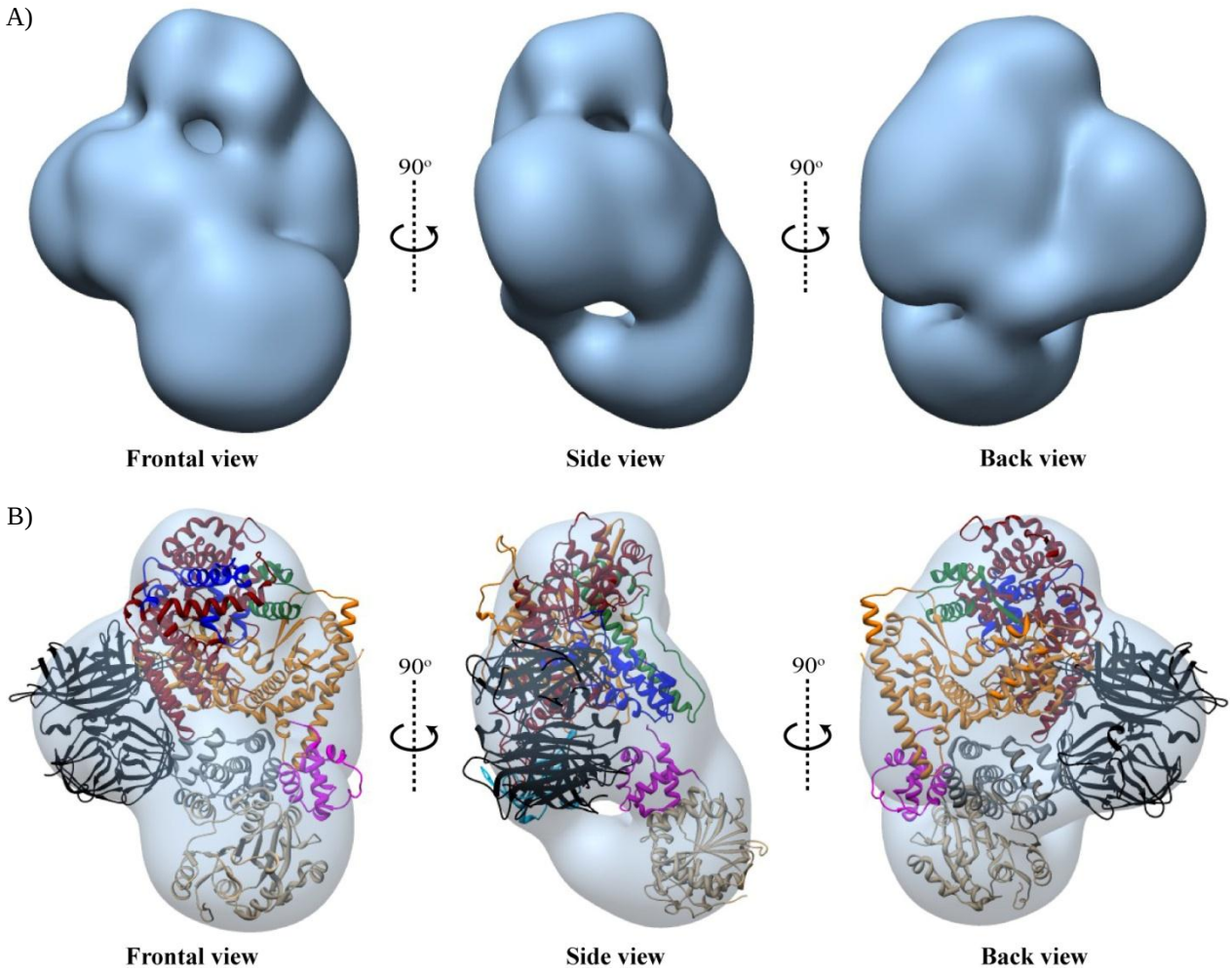


Figure 25: 3D reconstruction of downstream labeled IC. **A)** The final EM map of downstream labeled IC shown in three views: frontal, side and back. **B)** Fittings of atomic structures of human mtRNAP, Mtf1 and streptavidin. NTD, Thumb domain, Palm domain, Fingers domain and the additional N-terminal extension are colored in orange, green, red, blue and grey, respectively. The N-terminal α/β domain and C- terminal helical domain l of Mtf1, and streptavidin are dyed in tan, magenta and black. The comparison with upstream-labeled IC indicates a different position of the streptavidin tag. It seems that the downstream end of DNA seems to be placed deeper in binding cleft of sc-mtRNAP.

4.3. Structural characterization of EC

4.3.1. 3D reconstruction of negatively-stained EC

Huge conformational changes occur in T7 RNAP upon the transition from unstable initiation mode to highly stable and highly processive elongation mode (Yin y Steitz, 2002, Durniak *et al.*, 2008, Steitz, 2009). The increased stability can be also observed in the case of the artificial EC (Figures 26A and 26B, lanes marked as mtEC) that was formed by adding the sc-mtRNAP to previously annealed heteroduplex of template and non-template EC-DNA oligonucleotides and 10nt long RNA. The artificial EC are fully functional (data shown in the discussion) and thus represent the valid elongation state of the polymerase.

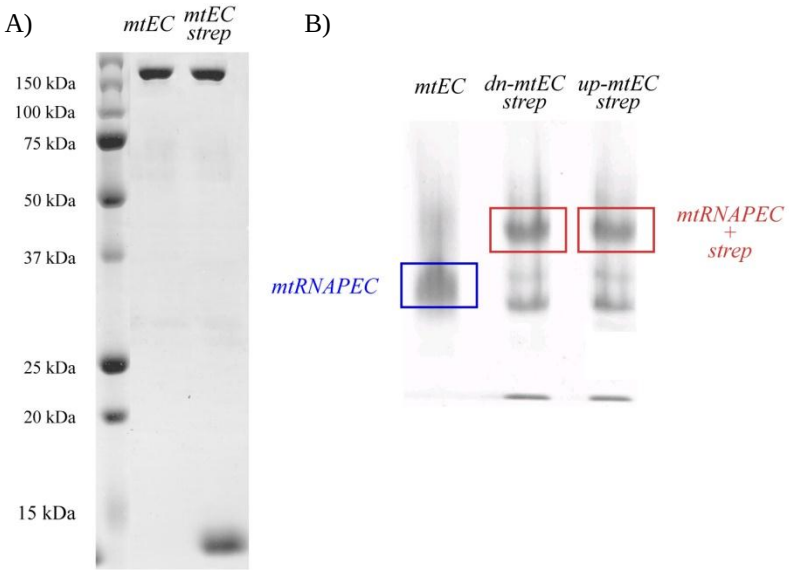


Figure 26: Electrophoretic analysis of mtRNAPEC and streptavidin labeled mtRNAPEC. A) 12% SDS gel: mtEC lane corresponds to mtRNAPEC, while lane marked as mtEC+step to downstream labeled complexes at 1:1 molar ratio **B)** 9% Native gel: mtEC lane corresponds to mtRNAPEC, while dn-mtEC and up-mtEC to downstream and upstream labeled mtRNAPEC, respectively. The labeled ECs (framed in red) exhibit a shift in migration due to the presence of streptavidin.

The previously mentioned rearrangements mostly affect the N-terminal portion, causing the breakage of its interactions with the promoter. The core subdomain, however, undergoes milder conformational changes exhibited through its 35Å translocation and 130^{o} rotation, without drastically affecting the internal structure of this portion. The tip of the fingers domain moves towards the downstream DNA, contributing the formation of the more extensive DNA binding site. The part of the thumb bends towards the fingers domain, which completes the formation of the pore that leads to the active site. The specificity loop and H subdomain, which arises through the reorganization of NTD, form the RNA exit channel. These changes are remarkably big for such a small protein, which implies that notable differences might be also detected between the IC and EC of mtRNAPs, even if a low resolution technique such as negative-staining EM is employed. With that aim we conducted a negative staining analysis of the mtRNAPEC. 20873 particles were selected and processed to obtain a final map of approximately 19 Å resolution. The 3D reconstruction revealed a less compact structure, approximately 20 Å wider and 10 Å deeper, comparing to quasi-IC (Figure

27). The frontal upper portion with the centrally positioned opening is more defined in negatively-stained EC than in quasi-IC, revealing the structure similar to the knuckle (Figure 27, red arrow).

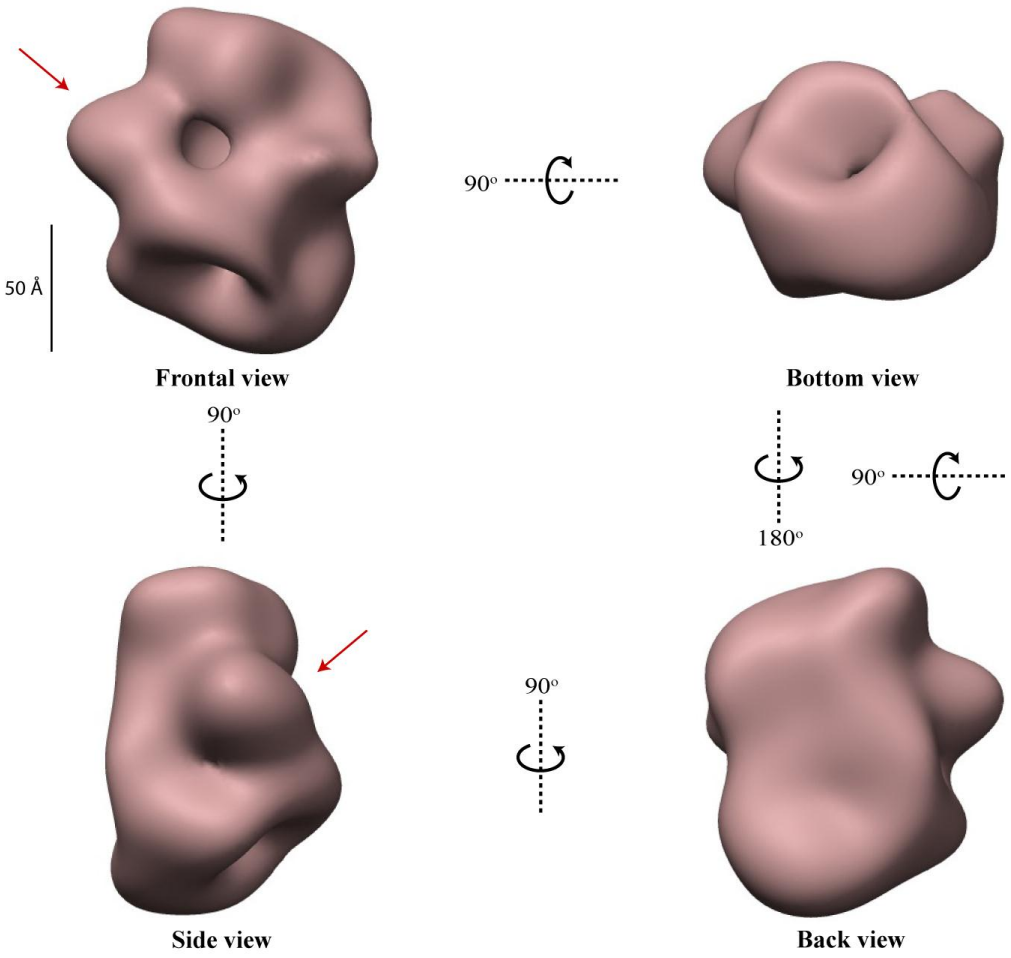


Figure 27: 3D reconstruction of EC. The final EM map is presented in four characteristic views: frontal, side, back and bottom. The red arrows (frontal and the side view) indicate towards the defined region that resembles to knuckle. The scale bar represents the length of 50 Å.

The atomic structure of human mtRNAP can be fitted fairly well in the EM map. The evident empty regions could be explained by the difference in the activity state of the two structures, the presence of the DNA and the difference in the sequences between the human and yeast mtRNAPs (Figure 28A). Firstly, sc-mtRNAP contains an insertion in the palm domain of ~90 amino acids, which function is still unknown. The secondary structure prediction of this region, preformed by Prof. Rui Sousa and his group, indicates the presence of a beta sheet and a loop organized to form the positively charged cleft, suggesting its possible role in the stabilization of the complex through the interaction with the negatively charged DNA (Figure 28B). Secondly, the part of palm and small portion of fingers domains might assume different positions in the EC, thus filling the region similar to the knuckle (Figure 27B, red encircled region). Thirdly, the DNA present in the EC might occupy the regions indicated by the black arrows. However, it should not be neglected the possible distortion of the observed particles that commonly occurs due to the surface tension forces created as the layer of stain dries. This is manifested through the flattening of spherical and cylindrical structures. Thus, these differences between the 3DEM map and atomic structure of human homologue could represent staining artifacts.

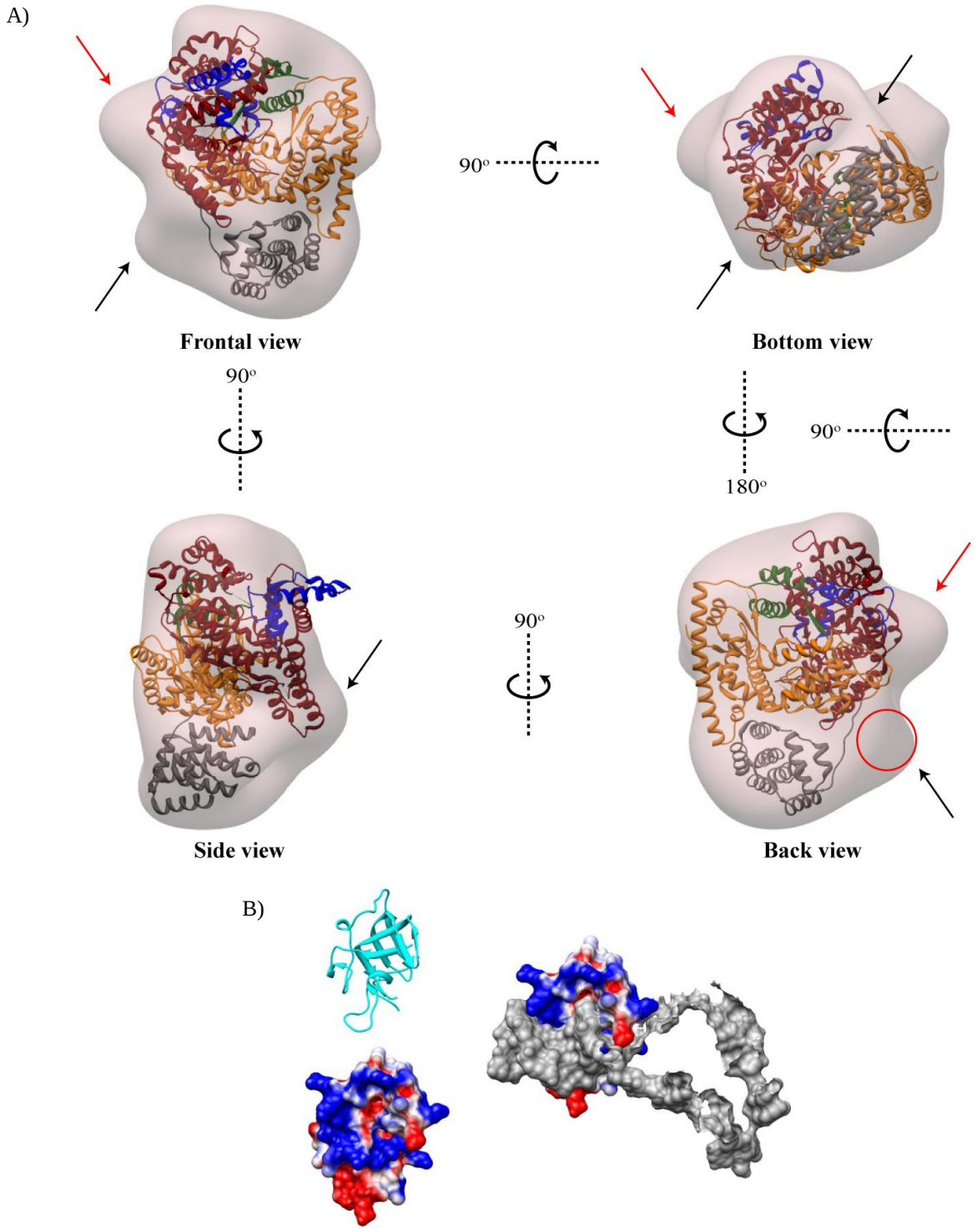


Figure 28: Docking of human mtRNAP X-ray structure into 3DEM map of negatively-stained EC.. A) The docking of human mtRNAP's atomic structure with color coded subdomains: NTD (orange); Thumb domain (dark green); Palm domain (dark red); Fingers (blue); and the additional N-terminal extension (grey). The docking is shown in four typical views: frontal, side, back and bottom. Red circle (back view) indicates the region that might be filed with ~90 amino acids long insertion of the palm domain. Arrows point to the remaining empty parts that might be occupied by the DNA molecule (black arrows), and the closely positioned alpha helices of the palm domain (red) and the small portion of the fingers domain (blue) that does not fit entirely into EM map (red arrow). **B)** Ribbon representation of the structure prediction of the palm subdomain insertion (cyan) and EC-DNA (template strand in red and non-template in green). The surface representation of the same structures, rendered in UCSF Chimera, is shown below, where blue represents positive, while red negative charges. The DNA surface is colored in grey

4.3.2. Cryo-EM of EC

The next logical step was to observe the EC under cryo-EM conditions, following the same preparation protocol as for IC (Header 3.3.2). 38210 particles were marked, boxed out, and subjected to 2D classification and subsequent 3D processing in EMAN and XMIPP. The independent reconstructions converged to a model that exhibited an overall shape similar to the model obtained for the negatively-stained EC, but was featureless despite that the individual particles exhibited remarkable signal to noise ratio, for such a small complex (~175 kDa) observed under cryo-EM. This can be an indicator of the existence of certain heterogeneity that could not be seen on the native gel (Figure 26). ML3D revealed the presence of an extreme flexible region (Figure 29, red arrows), that we named as the foot. This together with the small, but not insignificant population of particles without the DNA interfered with the earlier 3D reconstructions. The remaining ~11000 particles and the corresponding ML3D model were used to obtain the final 3DEM map at 18.9 Å. As mentioned before the cryo-EM density map originates from the unstained protein sample at almost physiological conditions, thus validating the results obtained by a negative-staining. The cryo-EM density map of EC has almost the same shape and the features as the one obtained from the negatively-stained sample, which clears any doubt that the observed features and protrusion are the staining artifacts (Figure 29).

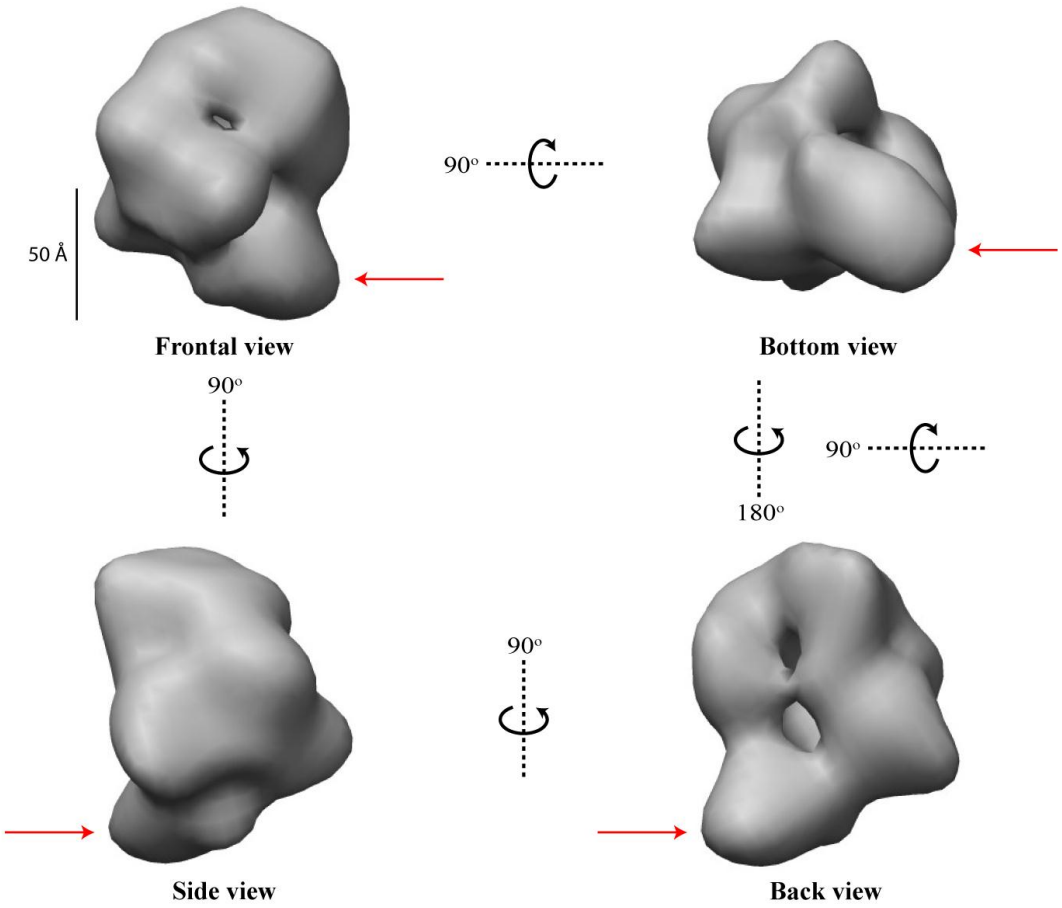


Figure 29: 3D reconstruction of EC obtained by cryo-EM. The final EM map is presented in four characteristic views: frontal, side, back and bottom. The red arrow point towards the region which has shown by ML3D to be quite mobile, and that has been named the foot. The scale bar represents the length of 50 Å.

For the fitting, the original atomic structure of human homologue was separated into 3 parts: C-terminal core with the thumb, the palm and the fingers subdomains; the NTD and the N-terminal extension, and the flexible loops that connect these regions. The N-terminal extension perfectly fits into the previously identified flexible foot. The more defined cryo-EM map allows the identification of extra regions, which can also be seen in the 3DEM map of negatively-stained EC (Figures 28B, black arrows), as the extremes of one separate element that further supports the assumption of being formed by the DNA (Figure 30, black arrows).

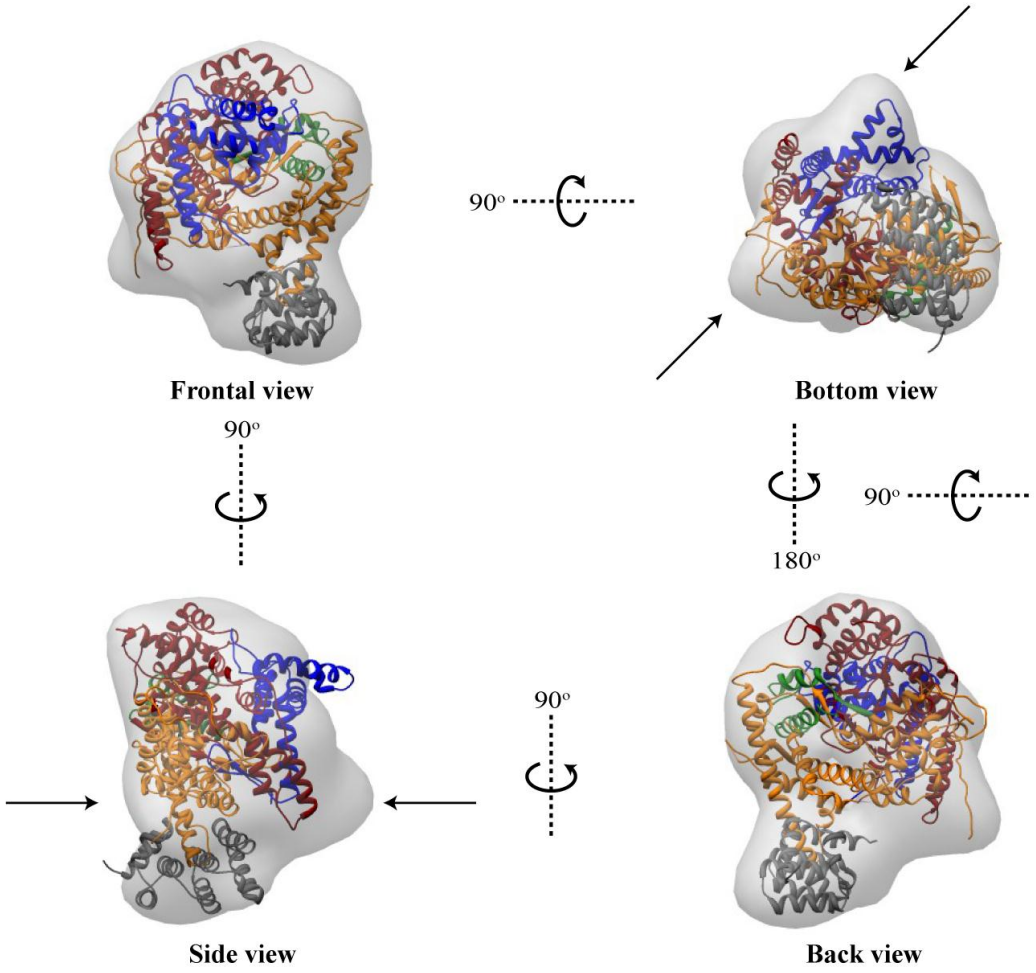


Figure 30: Docking of human mtRNAP X-ray structure into cryo-EM map of EC. The docking of human mtRNAP's atomic structure with color coded subdomains: NTD (orange); Thumb domain (dark green); Palm domain (dark red); Fingers (blue); and additional N-terminal extension (grey). The docking confirms the presence of the extra regions that were previously seen at the 3DEM model of negatively-stained EC, thus dispelling the doubts that might be artifacts of negative-staining. The docking is shown in four typical views: frontal, side, back and bottom. Arrows point to the remaining empty parts that might be occupied by the DNA molecule (black arrows)

4.3.3. Localization of upstream and downstream ends of DNA in EC

To address the question regarding the nature of an extra defined element, which presence is evident in 3DEM density maps of EC, we followed the same approach used for IC, which involves EM structural studies of upstream and downstream streptavidin-labeled ECs. The 3DEM maps would also help in answering the previously raised question regarding the potential existence of a common mechanism in T7RNAP and mtRNAPs for the transition from IC to EC and the promoter clearance. The high purity and

absence of degradation or contamination of EC and labeled EC was confirmed by SDS-PAGE. The native gels have indicated the presence of a smaller population of streptavdin-free ECs among the labeled ECs (Figures 26A and 26B, lanes marked as mtECstrep, dn-mtECstrep and up-mtECstrep). The 19886 and 13240 particles of downstream and upstream labeled EC, respectively, were used for subsequent 2D processing and 3D reconstructions. The mild heterogeneity was tackled at 2D and 3D level by repetitive ML2D, CL2D and ML3D classifications. As in the case of labeled quasi-IC and IC, the model of negatively-stained EC was filtered to low resolution and used as the initial model for ML3D. This allowed the segregation of the initial sets of particles to the ones that contained the information of streptavidin-bound ECs and the ones that corresponded only to non-labeled EC. These particles were afterwards classified by ML2D and CL2D. The resulting common class averages were used to generate the common lines initial model, which besides the generated blob, noise and low filtered non-labeled EC, was a starting point for independent iterative angular refinement in EMAN. Once the independently performed reconstructions converged to a common structure, the processing was passed to XMIPP (Projection Matching). The final model was filtered to a calculated resolution of 19 Å. At the 3DEM density map of upstream labeled EC, the extra mass, positioned at the right frontal part of the EC, can be noticed (Figure 31A and 31B). It protrudes from the area between the N-terminal extension, the NTD, the thumb and the portions of the palm and the fingers domains (Figure 31B), confirming our previous assumption regarding the DNA's composition of this region.

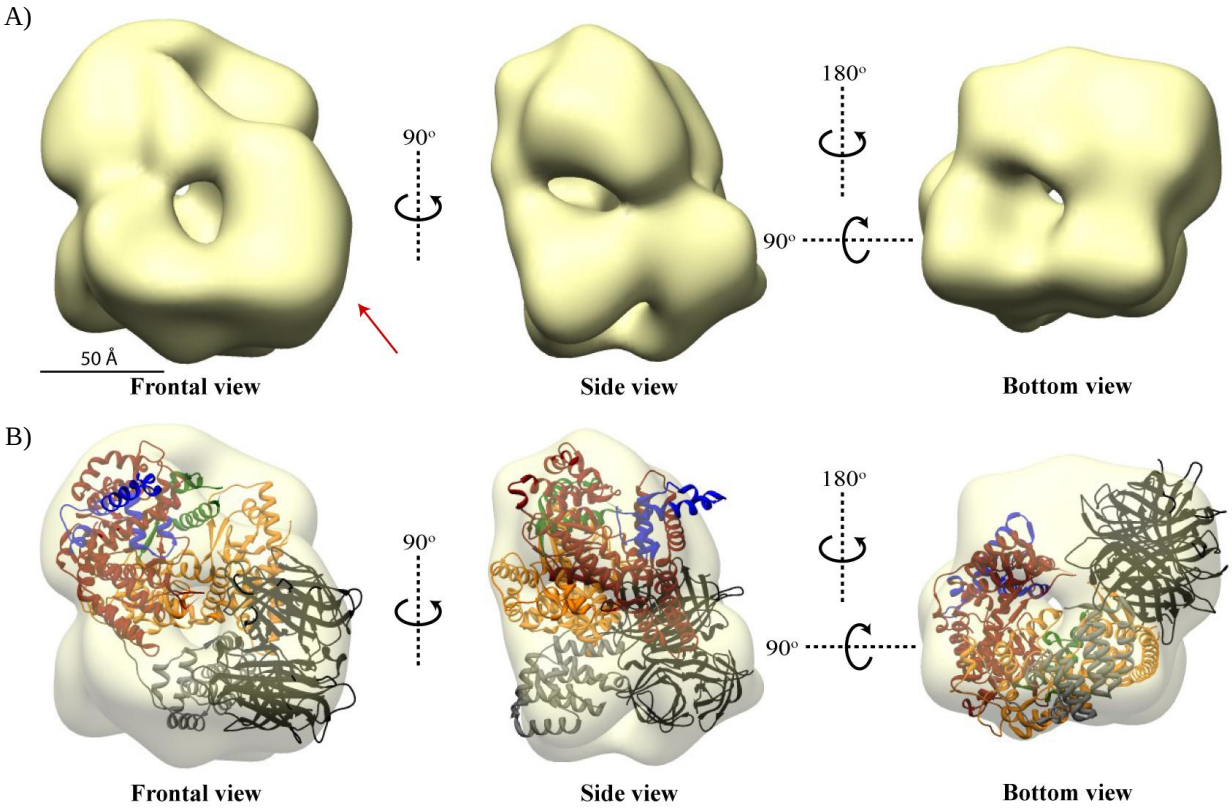


Figure 31: 3D reconstruction of upstream labeled EC. A) The final map originating from Projection Matching is shown in three characteristic views: frontal, side and bottom. The arrow indicates an extra mass connected to the frontal part of the sc-mtRNAP that corresponds to the bound streptavidin, indirectly identifying the position of the upstream end of the EC-DNA. The scale bar represents the length of 50 Å. **B)** Docking of human mtRNAP X-ray structure and streptavidin atomic structure into the 3DEM map of upstream labeled EC. NTD, Thumb, Palm and Fingers domains and the additional N-terminal extension are colored in orange, green, red, blue and grey, respectively. The atomic structure of streptavidin is dyed in black.

Further support to this is given by the 3DEM density map of downstream labeled EC that contains the extra mass placed at the other, left side of the electron-density that corresponds to EC (Figure 32A, red arrow). The observed extra protrusion at the back of the 3DEM map, which can't be seen at the 3DEM maps of non-labled EC, might be the displacement of the portion of NTD or even the C-terminal insertion element that is caused by steric clashes with the streptavidin (Figure 32B, black arrow).

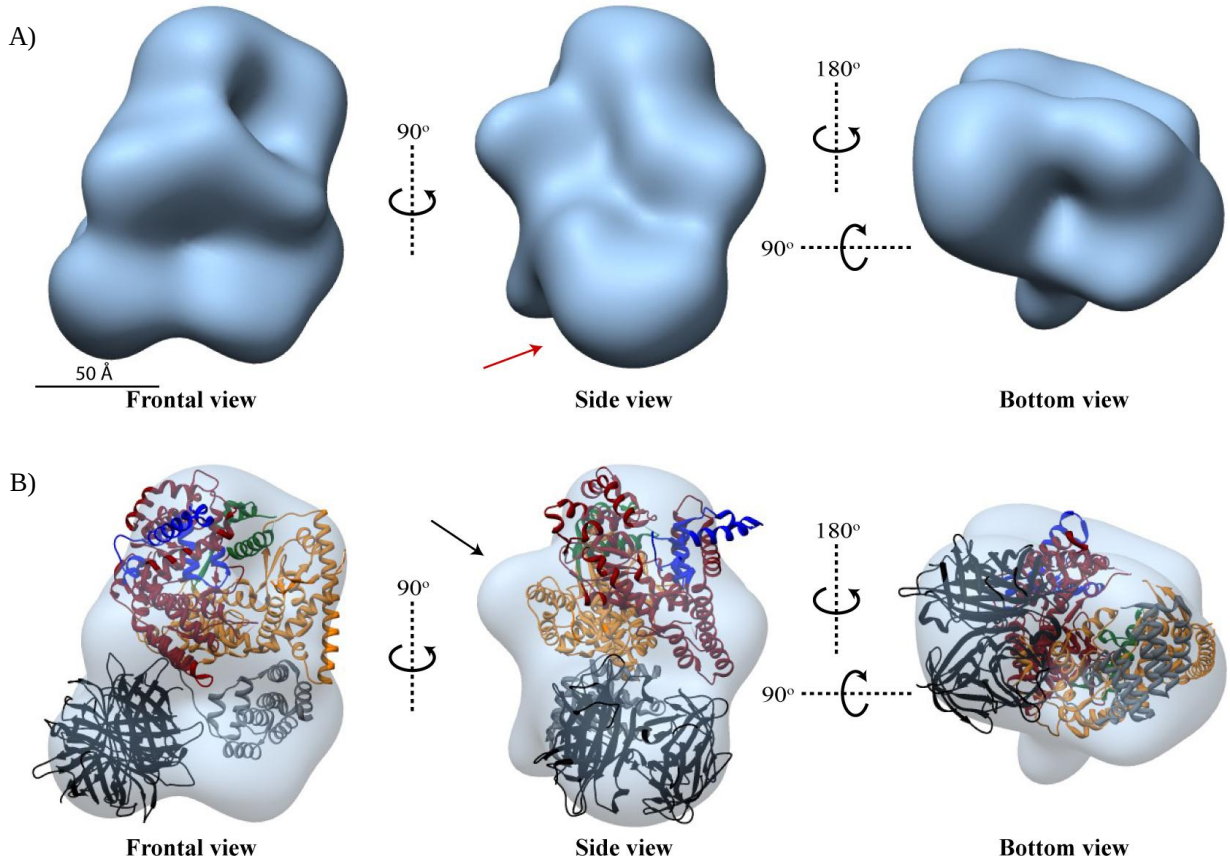


Figure 32: 3D reconstruction of downstream labeled EC. A) The final 3DEM density map of downstream streptavidin tagged EC is shown in three characteristic views: frontal, side and bottom. The model originates from 13240 selected particles and has the calculated resolution close to 19 Å. The red arrow (side view) indicates to an extra mass that corresponds to the bound streptavidin, indirectly identifying the position of the downstream end of the EC-DNA. The scale bar represents the length of 50 Å .**B)** Docking of human mtRNAP X-ray structure into the 3DEM map of downstream labeled EC. The subdomains of human mtRNAP are shown: NTD (orange); Thumb domain (dark green); Palm domain (dark red); Fingers (blue); and the additional N-terminal extension (grey). The atomic structure of streptavidin is dyed in black. The docking is shown in the same views as in A. The black arrow indicates to the extra mass that might be caused by the steric clashes with streptavidin. Notice the different position of the streptavidin mass with respect to upstream labeled EC (Figure 31).

4.4. Structural characterization of spRNAPIVEC

4.4.1. 3D reconstruction of negatively-stained spRNAPIVEC

The biochemical characterization of deletion mutants lacking the 1-100; 1-270 and 1-380 N-terminal residues have exposed multiple roles of sc-mtRNAP's N-terminal extension in the initial steps of mitochondrial transcription. Firstly, the residues between positions 100 and 380 mediate tight but dynamic interactions with Mtf1 prior the formation of ternary complex with the promoter, as the corresponding deletion mutants formed weak complexes with Mtf1. However, the presence of the promoter DNA might

induce the conformational changes in sc-mtRNAP and/or Mtf1 which expose additional regions for establishing the interaction between these two proteins, that results in a much stable and tighter complex. Secondly, the region spreading from the residue 270 to the residue 380 has been shown to be capable of formation of the upstream but not the downstream edge of the transcription bubble, thus identifying the importance of this region in the promoter opening. Thirdly, all three types of mutants have exhibited a lower affinity for ATP than the wt sc-mtRNAP, which implies that this region is involved in the binding of ATP and stabilization of RPo. Finally, the first 270 residues might have an auto regulatory role, as the average number of abortive events per one transcription cycle drastically dropped down in the case of N-terminal deletion mutant which lacked 270 residues (Tang *et al.*, 2009a). This might be the consequence of a looser interaction between the mutant and Mtf1 and its subsequent more efficient release, which is needed for the successful transition from IC to EC and release of the promoter. The small size of mtRNAP excluded the employment of immuno-EM in attempt to identify the localization of N-terminal extension in 3DEM maps of different activity states of sc-mtRNAP. One of the possible solutions was to generate a deletion mutant which missing part would be big enough to be registered through the comparison of 3DEM density maps of full length polymerase and its truncated version at the same activity state. The logical choice was the deletion mutant that lacked 380 residues from its N-terminal part, thus missing the entire N-terminal extension (first 300 residues). Furthermore, solving its structure would also give the insight in the architecture of the alternative splicing product of human mtRNAP gene that has been shown to be involved in transcription of protein-encoding genes in the nucleus of mammalian cells, besides the RNAPII (Kravchenko *et al.*, 2005). Therefore, we intended to use a truncated form of sc-mtRNAP to generate quasi-IC, IC and EC. Unfortunately, all the formed complexes, with the exception of the EC, exhibited severe aggregation and thus were unfit for EM studies (data not shown). The formed truncated EC showed the dissent stability at the native gels (Figure 33).

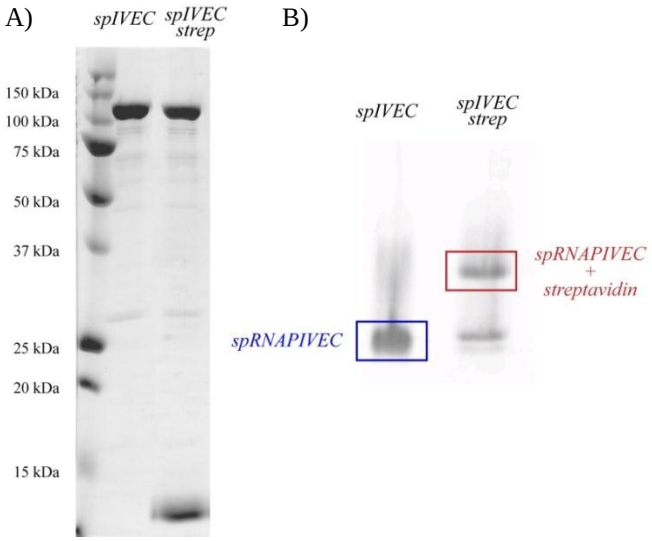


Figure 33: Electrophoretic analysis of spRNAPIVEC and streptavidin labeled spRNAPIVEC. A) 12% SDS gel: the lane marked as spIVEC corresponds to spIVRNAPEC, spIVECstrep to downstream labeled spIVRNAPEC at equimolar ratio **B)** 9% Native gel: spIV lane corresponds to spRNAPIVEC, while lane marked as spIV+step. to downstream labeled complexes. The downstream labeled ICs (framed in red) exhibit a shift in migration due to the presence of streptavidin.

17258 particles of the remaining spRNAPIVEC complex were submitted to 2D and 3D image processing, resulting in a 3DEM model of similar size and shape as the EM map of the full length EC. The only notable difference is at the left side, where in the case of spRNAPIVEC the electro-density is missing (Figure 34, red arrow). Thus, an extra opening to the centrally positioned cavity can be seen. The missing electro density matches with the proposed position of the DNA in both the cryo and the ns-EM map of the full length EC (Figure 34). This differential staining can be the result of combined effect of several factors: Firstly, it is known that protein:nucleic acids complex might be severely affected by exposing the protein:nucleic acid complex to low pH conditions such as in the case of staining with uranyl acetate, a heavy metal die that operates at pH close to 3.4 (Aramayo *et al.*, 2011). Secondly, the absence of the N-terminal extension might cause that the DNA is bound less tightly to spIVRNAP and thus being more sensitive to uranyl acetate staining.

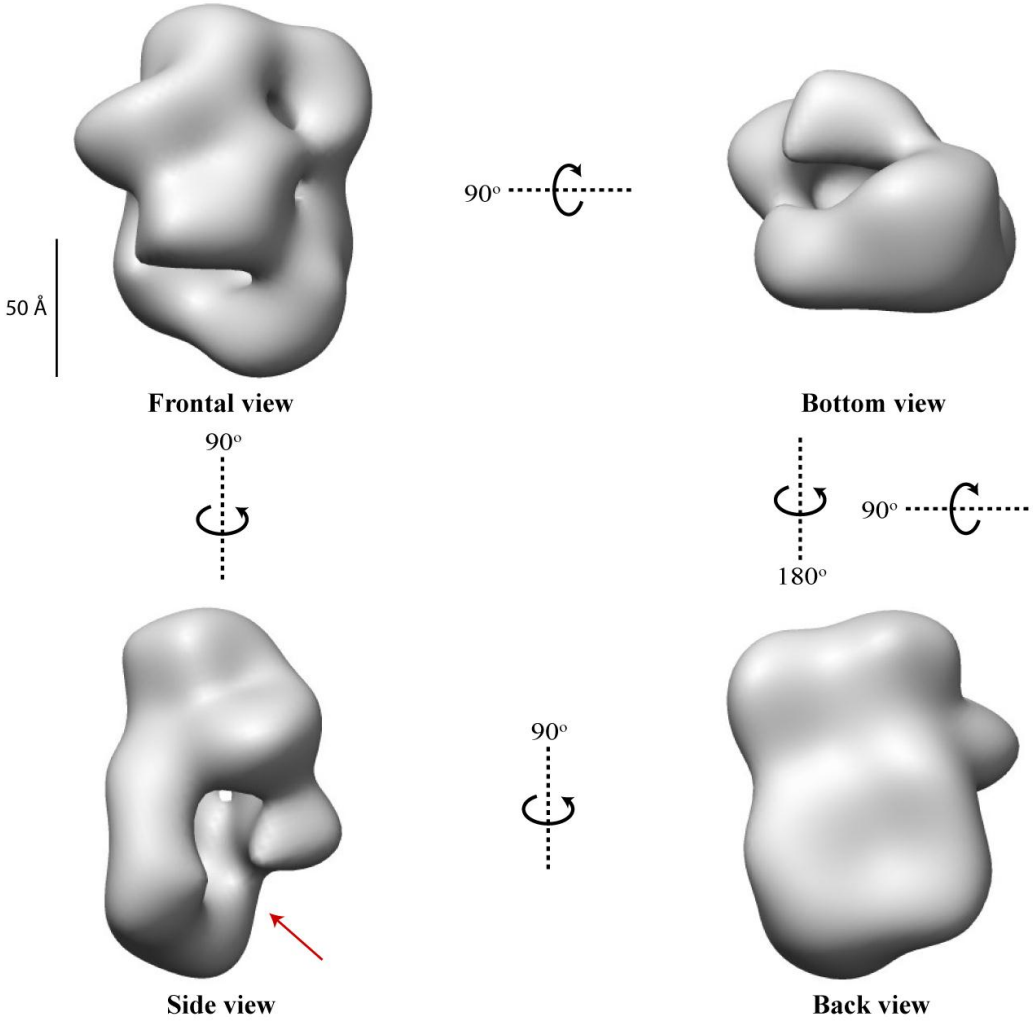


Figure 34: 3D reconstruction of spRNAPIVEC. The obtained EM map is presented in four characteristic views: frontal, side, back and bottom. The missing electron density, which is best seen at the side view (indicated by the red arrow), is probably due to the uneven staining of the DNA in the two cases (spRNAPIV and mtRNAP ECs). The scale bar represents the length of 50 Å

As in the case of the mtRNAPEC, the docking of the atomic structure of human mtRNAP provides further validation of the final model. The non-filled portion, composed of the downstream EC-DNA, can also be seen (Figure 35).

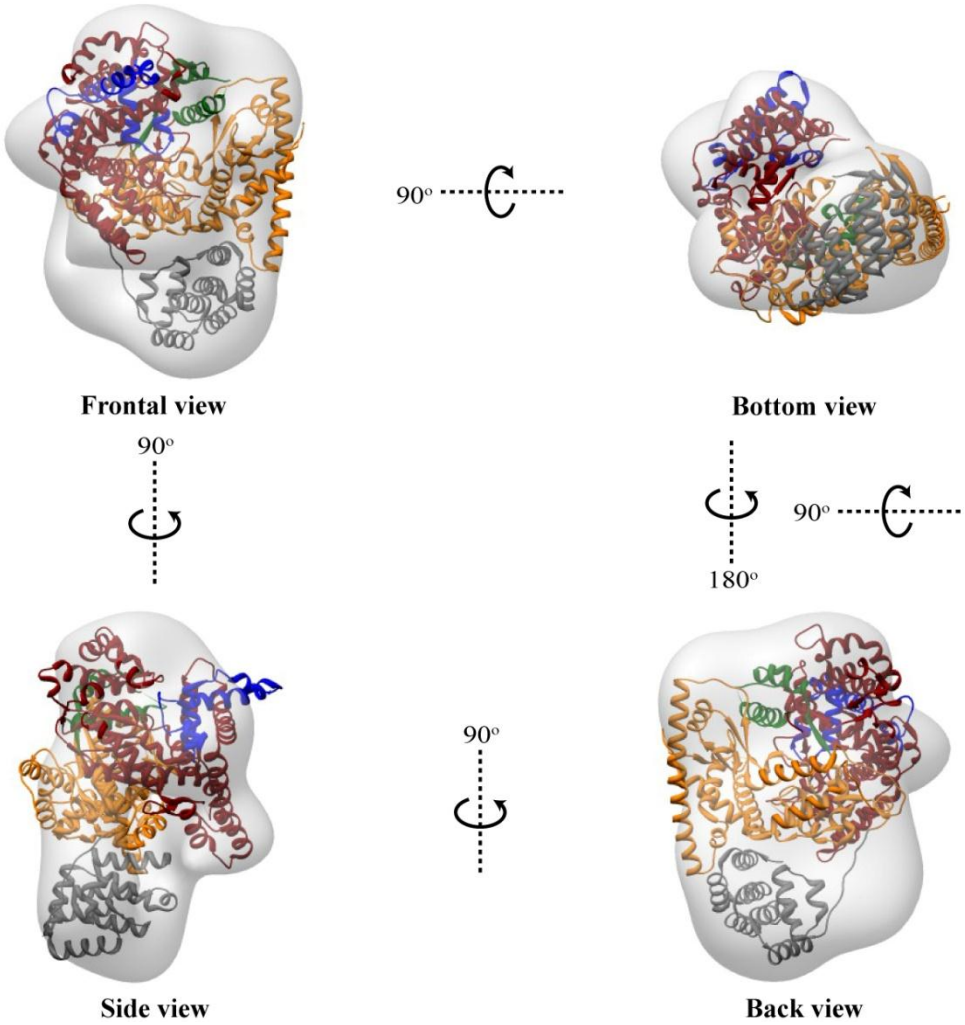


Figure 35: Docking of atomic structure of human mtRNAP into 3DEM map of spRNAPIVEC . The subdomains of human mtRNAP are shown: NTD (orange); Thumb domain (dark green); Palm domain (dark red); Fingers (blue); and the additional N-terminal extension (grey).

4.4.2. Localization of downstream end of DNA in spRNAPIVEC

The structural characterization of spRNAPIV would help in addressing the question if the truncated protein exhibits different interaction with the DNA, which could be manifested through its altered position comparing to the complex with full length protein. As previously mentioned, the extreme tendency of truncated sc-mtRNAP for aggregates formation has made only the spRNAPIVEC suitable for EM. The purity and stability of the downstream labeled spRNAPIVEC were assessed by SDS-PAGE and Native-page, which showed the existence of small but notable population of non-tagged spRNAPIVEC (Figures 32A and 32B, lanes marked as spIVECstrep). 23848 particles of downstream streptavidin tagged spRNAPIVEC were selected, 2D classified and submitted to 3D processing. Unlike expected, the resulting 3DEM model (Figure 36A, arrow) contains the mass corresponding to streptavdin almost at identical position as in the 3DEM density map of downstream labeled mtRNAPEC (Figures 31 and 32). Also the model lacks even more mass in the left frontal part that has been identified to be contributed by the DNA (Figures 27-30). The synergistic effect of unstable staining of DNA and the possible acquisition of a defined structure/position of the rest of DNA through the interaction with the N-terminal extension, might explain the difference between the

mtRNAPEC and spRNAPIVEC in this region. Thus, the model of streptavdin tagged spRNAPIVEC provides an additional verification of the 3DEM models of full length EC, streptavdin tagged EC, and spRNAPIVEC. We performed docking of known X-ray structures for human mtRNAP and streptavidin using UCSF Chimera. To facilitate the manual fitting of crystallographic structures into portions of the map, the downstream labeled map of spIVRNAPEC was segmented using the segmentation option in UCSF Chimera. The manual fittings were then rendered using the Fit-to-Map command in Chimera which is based on the determination of real time correlation between the reference model and the surface created from the provided atomic structure/model (Figure 36B).

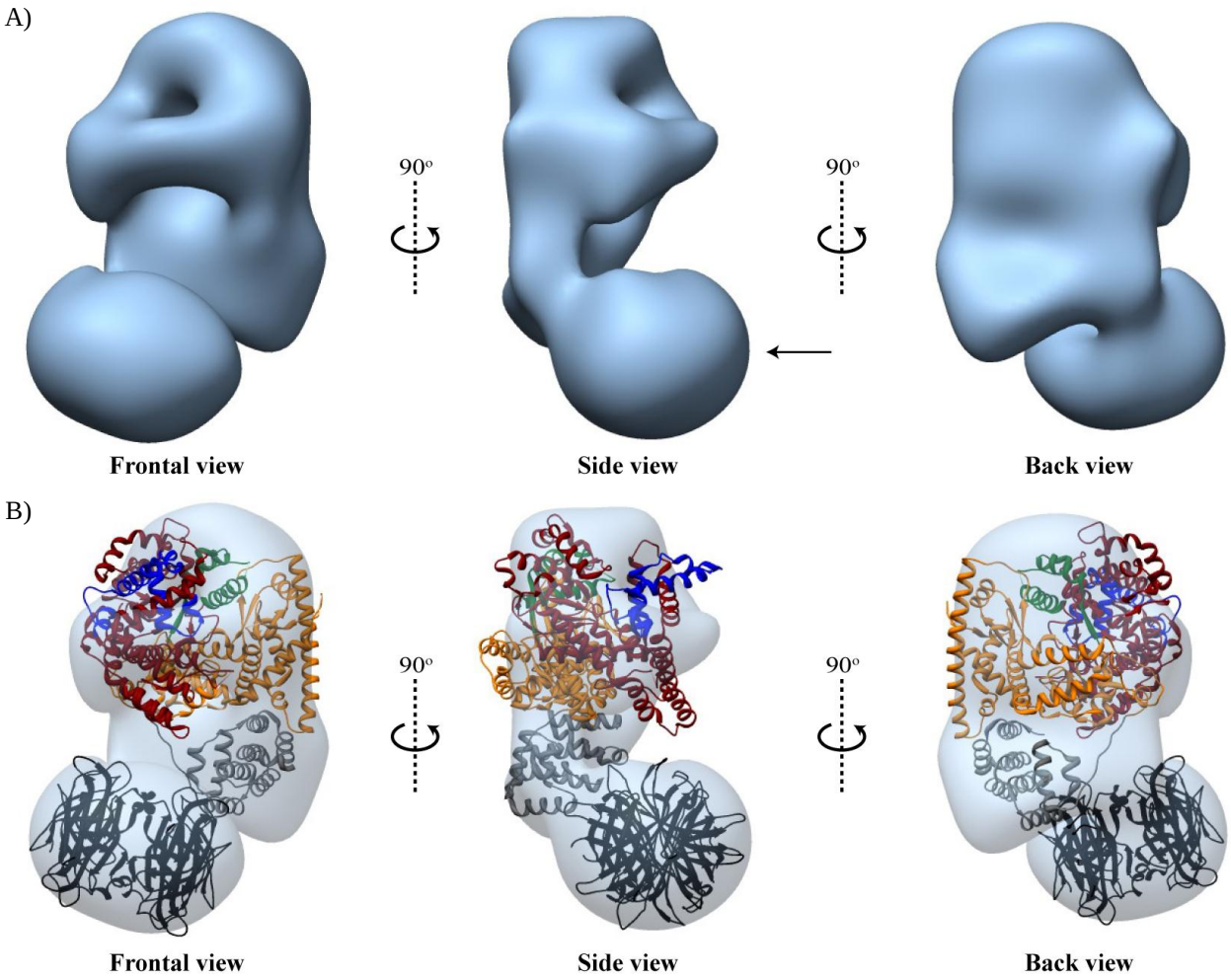


Figure 36: 3D reconstruction of downstream labeled spIVRNAPEC. A) The final map originating from Projection Matching is shown in three characteristic views: frontal, side and bottom. The black arrow (side view) indicates an extra mass connected to the frontal part of the sc-mtRNAP that corresponds to the bound streptavidin. The scale bar represents the length of 50 Å **B)** Docking of human mtRNAP X-ray structure into downstream labeled spIVRNAPEC. The subdomains of human mtRNAP are shown: NTD (orange); Thumb domain (dark green); Palm domain (dark red); Fingers (blue); and the additional N-terminal extension (grey). The atomic structure of streptavidin is dyed in black. The docking is shown in the same views as A. Notice the same position of streptavidin as for the downstream labeled mtRNAPEC (Figures 31 and 32).

5. Discussion

5.1. Promoter recognition, binding and melting

The minimal transcription machinery in yeast mitochondria required for the normal level of expression of its genes, consists of sc-mtRNAP (Rpo41) and Mtf1 (Mangus *et al.*, 1994). The average yeast cell contains approximately 50 molecules of Rpo41, 20 molecules of Mtf1 and 10-30 copies of mitochondrial genome (Grimes *et al.*, 1974, Mangus *et al.*, 1994). This has two implications: (1) there is nearly one molecule of holoenzyme per one copy of mitochondrial genome, and (2) the amount of the present Mtf1 is a limiting factor in mitochondrial gene expression. The identification of T7 RNAP's three key elements, responsible for promoter recognition, binding and melting (AT rich recognition loop, intercalating β hairpin and specificity loop) in sc-mtRNAP raises the questions: why the mtRNAPs require the presence of auxiliary factors and what would be their function in transcription initiation. One of the reasons may reside in the different structure of the mitochondrial promoters compared to the ones recognized by T7 RNAP, which implies a different geometry of the RNAP:promoter interaction and the possible existence of alternative or additional mechanisms, employed in promoter recognition, binding and melting (Savkina *et al.*, 2010) (Figure 37).

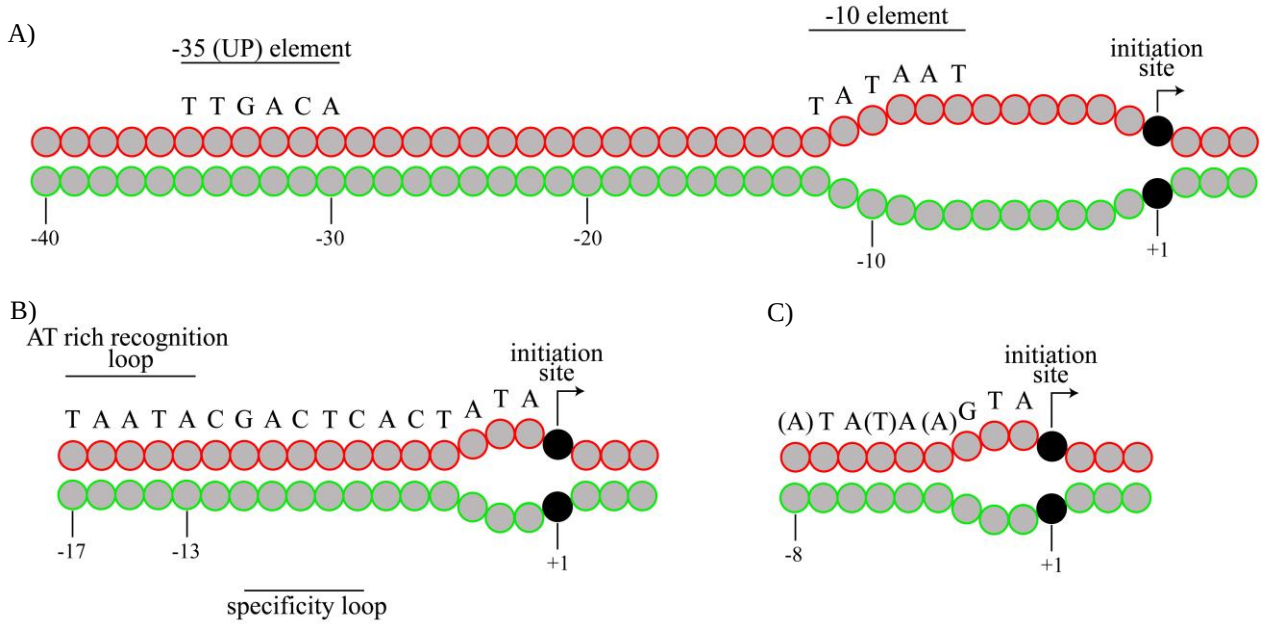


Figure 37: Promoter organization in *E.coli*, T7 phage and yeast's mitochondria with the transcription bubble. A) Promoter motifs recognized by primary bacterial RNAP s factors. Gray circles represent the DNA nucleotides (top with red rim, non-template strand; bottom with green rim, template strand). The transcription bubble in RPo, which spreads from -12 to +2, is illustrated (separated circles). Promoter motifs recognized by σ factors are marked: UP or -35 element that is recognized by the σ4.2 region of σ4 subdomain, and the -10 element recognized by the σ2.3 and σ2.4 regions. The circle dyed in black represents the initiation site. **B)** Class III T7 phage promoter. Motifs that interact with the AT rich recognition loop and the specificity loop are emphasized. The transcription bubble extends from the positions -4 to +2 (separated circles). The addition of the first nucleotide to RNA is followed by the additional unwinding to the position +3 or +4. The color code of the circles is same as in A) **C)** The yeast mitochondrial promoter. The nucleotides in brackets are the non-conserved ones. The transcription bubble has similar size as in the case of class III T7 promoters (separated circles). However, it is not expanded upon the addition of the first NTP. The color code of the circles is same as in A)

The high homology with T7 RNAP (Masters *et al.*, 1987), which indicates the phage origin of sc-mtRNAPs and of the mtRNAPs in general, and the requirement for the presence of auxiliary factors,

positions the mitochondrial transcription machinery as the evolutionary link between the single and multi-subunit RNAPs. The transcription in bacterial systems begins with the formation of the functional holoenzyme, as neither the core nor σ factor can bind to DNA by themselves. The conformational changes induced by the core in the σ factor result in the displacement of negatively charged σ1.1 subdomain from the σ2 and σ4 subdomains (Schwartz *et al.*, 2008). Once released from σ1.1 subdomain's inhibitory effect, the σ2 and σ4 can establish sequence specific interactions with the DNA. The binding of σ4 to the region that encompasses nucleotides -38 to -30 induces the bending of the UP (^{35}TTGACA^{30}) element, which brings it in close proximity of αCTD, permitting their interaction (Bose *et al.*, 2008). At the same time, the σ2.4 region establishes contacts with the upstream end of -10 element (^{12}TATAAT^{7}) (Murakami *et al.*, 2003). All these events allow the formation of RPc (promoter closed complex) in which the holoenzyme and its elements, including the σ2.3 region, acquire an adequate position and a correct orientation towards the DNA. The AT-rich content of the -10 element is responsible for its natural breathing and a frequent single strand exposure. The previously mentioned binding events and the induced structural changes of the promoter allow the perfect positioning of the σ2.3 region in close vicinity of the -10 element and the subsequent stabilization of the exposed single strand. As a consequence of this, the upstream edge of the transcription bubble is formed (Mekler *et al.*, 2002, Murakami *et al.*, 2002b). The unwinding goes until the bubble reaches the nucleotide at position +2 (downstream edge). The key step during this process is the initial unwinding from -11 to -7 that would introduce flexibility in the DNA and lead to the entering of approximately 20 bps long downstream DNA duplex into the DNA-binding clamp (Borukhov y Nudler, 2008). All these events are necessary for the correct placement of +1 base of the template strand in the catalytic center.

As previously mentioned, T7 RNAP is capable of binding DNA by itself, and three structural elements are responsible of this: the AT rich recognition loop, the specificity loop and the intercalating β hairpin. The AT rich recognition loop recognizes and binds to the sequence from position -17 to -13 of T7 class III promoter's binding domain. This induces a slight bend in the axis of the DNA helix, making the region between -11 and -7 accessible to the specificity loop (Cheetham *et al.*, 1999). The specificity loop establishes the contacts with nucleotides at positions -11 and -10 of non-template strand and the region that spreads from -9 to -5 of the template strand (Rong *et al.*, 1998). The initial binding of the AT-rich recognition loop and the specificity loop causes the DNA to bend 40 ^{o}-60^{o} around its -1 bp. This leads to accumulation of the tension and the subsequent opening of the double stranded DNA in the region with the lowest energy content, the TATA region of the initiation domain. The previous bindings also allow the intercalating β hairpin to be properly placed in the proximity of the initiation domain. The intercalating β hairpin gets inserted between the nucleotides at position -4 and -5, which stabilizes the upstream edge of the transcription bubble and separates the template from the non-template strand. The template strand is further directed to the catalytic active site and its 5`end is placed into the pocket formed in the junction of the palm, the fingers and the N-terminal domains (Cheetham *et al.*, 1999, Cheetham y Steitz, 1999). The aromatic side chain of Trp422 stacks on the -1 base, inducing a sharp bend in the template strand and positioning the base at the position +1 as the initial one. The incorporation of the GTP as the first nucleotide stabilizes the RPo and leads the

additional unwinding of the DNA for 1 or 2 nucleotides. At that moment, the transcription bubble spreads to the positions +3 or +4, thus being 7 or 8 bases long.

The sc-mtRNAP exhibits a relatively high affinity for DNA, regardless of the DNA content. This means that sc-mtRNAP does not contain the mechanism to distinguish the promoter DNA sequences from the non-promoter ones at the initial DNA binding step (the equilibrium dissociation constant (Kd) values are in the same range; i.e. 48-66 nM). On the other hand, the presence of Mtf1 drastically lowers the dissociation rate of sc-mtRNAP from DNA, without affecting their association rates. The stabilization effect of Mtf1 is more pronounced on the sc-mtRNAP:promoter-DNA complexes than on the ones with non-promoter DNA, resulting in 4 times higher affinity of sc-mtRNAP:Mtf1 complex towards the promoter DNAs (Tang *et al.*, 2009a). However, this difference is negligible compared to the T7 RNAP, which exhibits 10^5 higher affinity for the promoter than for the non-promoter sequence (Bandwar y Patel, 2002). This implies the use of a different mechanism that allows the sc-mtRNAP:Mtf1 complex to discriminate between the promoter and the non-promoter sequences. It seems that this new mechanism is based on the differential bending of DNA, which in the presence of Mtf1 results in a change of bending angle of the promoter DNA from 52^o to 89^o, which cannot be observed for non-promoter DNAs. Furthermore, only the DNAs that are bent for 89^o are subsequently melted, which means that the bending and melting are coupled processes (Deshpande y Patel, 2012). It is important to notice that a similar change in the bending angle of promoter DNA is observed for T7 RNAP. Namely, the interactions between the AT rich recognition loop and specificity loop with the DNA sequences from position -17 to -13 and from -11 to -7, respectively, cause the DNA to bend for 40^o-60^o. The incorporation of the 8th nucleotide in the nascent RNA molecule is followed by the left handed rotation of the PBD and the specificity loop by 45^o (Steitz, 2009). As the DNA remains bound to the enzyme, it undergoes an additional 30^o-35^o bending (Durniak *et al.*, 2008). Our structural data is in a full agreement with these results. The mtRNAP and promoter DNA can interact, as described above, but the complex is not sufficiently stable to be structurally characterized by EM. Instead, a stable and transcriptionally active complex can be formed with the bubble promoter, even in the absence of Mtf1 and the presence of the substrate (Matsunaga y Jaehning, 2004, Tang *et al.*, 2009a). The comparison between the quasi-IC and IC volumes generated reveal two extra electron-densities: (1) a large one present at the front of sc-mtRNAP and that corresponds to Mtf1; (2) a smaller one visible as the extra region that protrudes from the back of sc-mtRNAP (Figure 38). The subsequent labeling of upstream DNA of IC confirms the assumption that this extra region is partially composed of DNA (Figure 24). Furthermore, the comparison between the 3DEM maps of upstream labeled quasi-IC and IC shows a significant displacement of upstream end of DNA in two states, thus being in accordance with the previously mentioned differential bending of promoter DNA by sc-mtRNAP alone and sc-mtRNAP:Mtf1 complex.

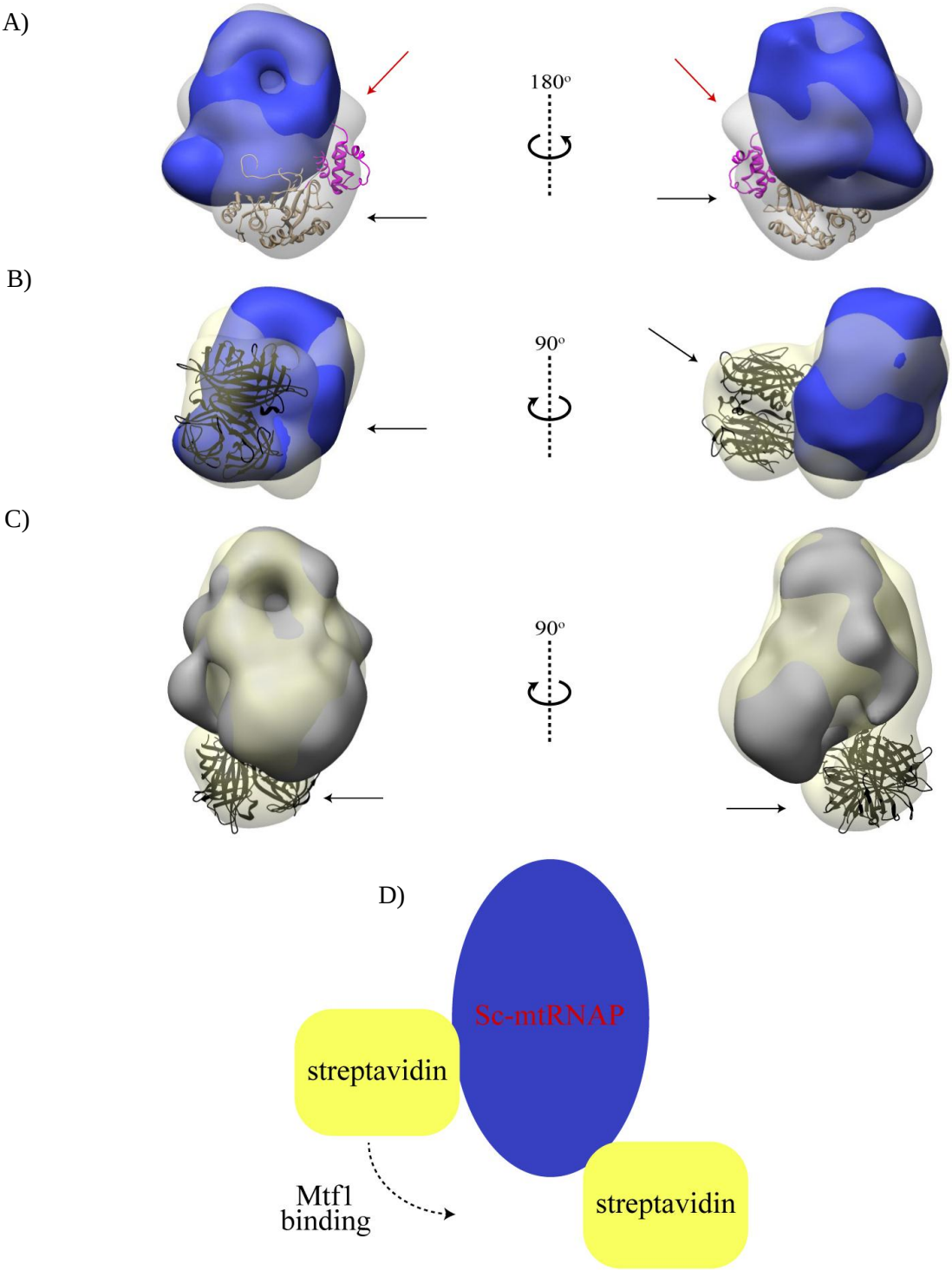


Figure 38: Differential position of the upstream DNA in the quasi-IC and IC. A) Comparison of the 3DEM density maps of quasi-IC (blue) and IC (transparent grey). The black arrows point to the Mtf1 electron density, with the atomic structure of Mtf1 docked in, and the electron density formed from the upstream DNA and C-terminal tail (red arrow). Two superpositioned EM maps are shown in the frontal view (left) and the view that is obtained by 180 ° rotation. **B)** Comparison between the 3DEM density maps of quasi-IC (blue) and upstream streptavidin tagged quasi-IC (yellow). The mass that corresponds to streptavidin is indicated by the black arrow and contains the docked streptavidins atomic structure (black). **C)** Superposition of 3DEM density maps of IC and upstream streptavidin labeled IC revealing the extra density (black arrow) that originates from streptavidin. **D)** Schematic illustration of dislocation of the upstream DNA (indirectly identified by the biotin-streptavidin labeling) that occurs upon the binding of Mtf1. This result is in the agreement with the findings of Deshpande and Patel (2012).

The sc-mtRNAP:Mtf1 complex melts the promoter from -4 to +2, thus the initial transcription bubble has a similar size both in yeast and in T7 page . In the case of yeast mitochondrial transcription, the addition

of the first nucleotide in the RNA is followed by a slight structural rearrangement of the region that consists of the nucleotides at the positions -2 and -1. However, this does not provoke any change in the downstream edge of the transcription bubble, which remains at the +2 position (Tang *et al.*, 2009a). The Mtf1 in the RPc and RPo seems to establish contacts with the region that extends from the positions -10 to -8, which suggests its involvement in the recognition of the upstream promoter elements (Paratkar y Patel, 2010, Savkina *et al.*, 2010). Furthermore, Mtf1 interacts with nucleotides from position -5 to -2 and from +1 to +2 of the non-template strand, and with the nucleotides at +4 and +5 of the template strand (Paratkar *et al.*, 2011). Seemingly confronting results (Savkina *et al.*, 2010) indicate the interaction of the C-terminal tail of the Mtf1 (residues 320-341) with the nucleotides at positions -4 and –3 of the template strand. These results seem contradictory, as the transcription bubble starts from at position -4. Therefore, the C-terminal probably plays a role in the formation of the upstream edge of the transcription bubble. This may occur through the stabilization of single strand that gets exposed as a consequence of the DNA bending and accumulation of stress. The interaction between the C-terminal tail of Mtf1 and the nucleotides at the positions -4 and -3 suggest that this portion of Mtf1 becomes buried in the active binding cleft of sc-mtRNAP. Our 3DEM reconstructions of IC and the chemical nuclease experiments performed by the group of our collaborator Prof. Rui Sousa, which were based on our structural IC 3DEM volumes and subsequent docking, confirm this assumption. Namely, the residues indicated by the fitting (K324, V332 and S340) were substituted by cysteins, which were subsequently conjugated with Fe (III) (S)-1-(p-Bromoacetamido-benzyl)ethylene diamine tetraacetic acid (Fe-BABE) nuclease by incubation for 1h at 37^{o}. Unbound Fe-BABE was removed by gel filtration, while the nuclease tethered mutants were used for the cleavage reactions. The nucleases of tethered ICs, which were halted at +0, +3, or +7,(RNA length in nts), were activated by adding the ascorbic acid and hydrogen peroxide (H_2O_2). The reactions were stopped after 5s by addition of the stop buffer (95% formamide, 20mM EDTA, 0.1% xylem cyanol). The reactions products were analyzed by electrophoresis (Figure 39A). Furthermore, the experiments with carboxypeptidase Y show that the C-terminus of Mtf1 alone or Mtf1+DNA is rapidly digested, with approximately 8 residues being rapidly removed from the C-terminus (mass spec data, not shown). However, when the same is repeated with IC, the C-terminus is strongly protected from digestion (Figure 39B). This indicates that the C-terminus becomes buried when Mtf1 binds the IC, and the chemical nuclease and functional studies of Mtf1 with a C-terminal deletion suggest that the C-terminus actually “invades” the polymerase template binding cleft and DNA bubble. The remaining interactions with the non-template strand up to nucleotide at position +2 may play a role in the further stabilization of RPo and in promoter unwinding. This suggests that Mtf1 has a similar function as the σ factor of bacterial holoenzyme. The close proximity of Mtf1 to the transcription site and the tighter binding of sc-mtRNAP and the first NTP incorporated in the RNA molecule in its presence, indicate that the Mtf1 plays a role in the binding of first two NTPs (Tang *et al.*, 2009a, Paratkar *et al.*, 2011).

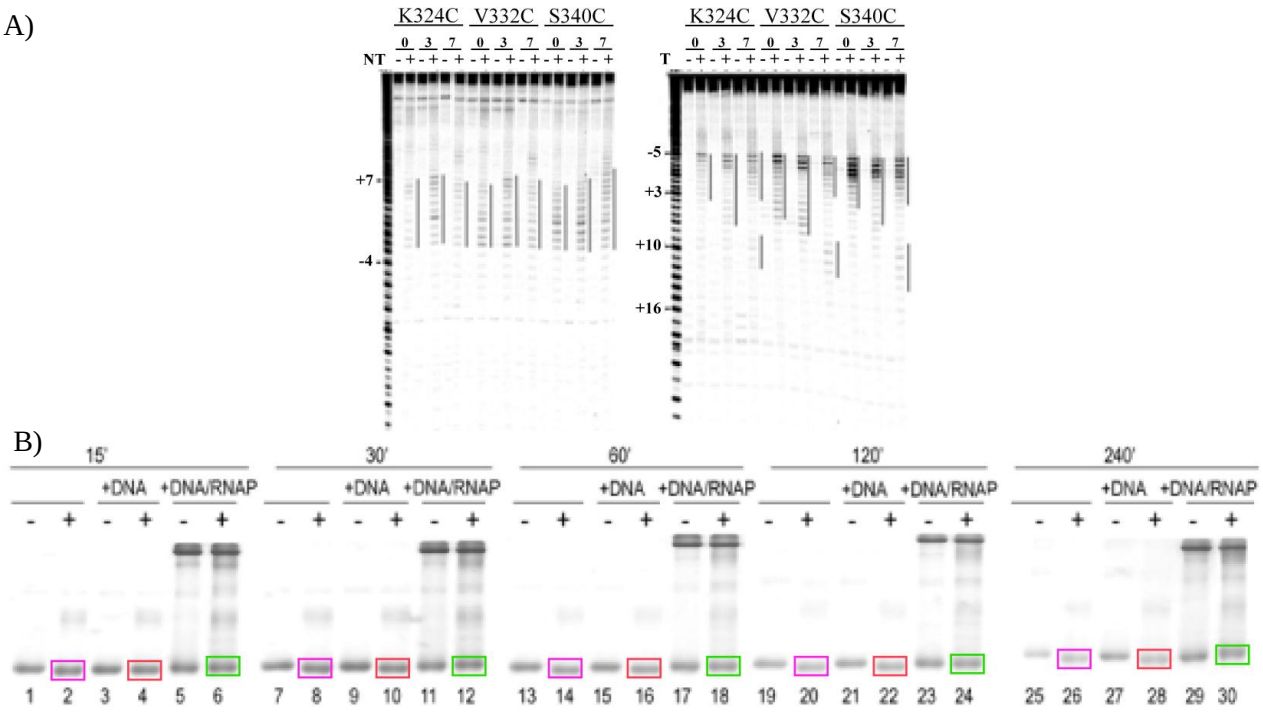


Figure 39: Fe-BABE nuclease and carboxypeptidase Y digestion pattern. A) The chemical nuclease cleavage pattern when nuclease is tethered to residues at position 324, 332 and 340. Lanes labeled with 0,3 and 7 correspond to digestion product when no NTPs (RNA=0 nts), ATP+3'dUTP (RNA=3 nts) and ATP+UTP+3'dGTP were added (RNA=7 nts). The changes in cutting therefore show the polymerase moving and changing conformation on the DNA as it progresses through the transcription initiation reaction. On the T-strand that, as the RNA is extended from 0 to 3 to 7 nts (left to right) in each set of lanes that the upstream cutting patterns (between -2 and -5) remain in the same place, but the downstream patterns (from +3 to +16) move downstream as the RNA is extended. This indicates as the RNA is extended from 0 to ~7 nts, the RNAP maintains its interaction with the upstream promoter. However, as the RNA is extended the downstream DNA does have to move through the RNA and is accommodated by a combination of "scrunching" and conformational changes in the RNAP. The movement of the downstream DNA is evident as the changes in the cutting sites. The "-" and "+" lanes correspond to reactions in which peroxide was either added (+) or not added (-). **B)** Addition of carboxypeptidase Y to Mtf1 alone (lanes 2, 8, 14, 20, and 26; purple frames) or MTF1+Bubble promoter DNA (lanes 4, 10, 16, 22, and 28; red frames) results in the rapid removal of 8 residues from the Mtf1 C-terminus as seen by the shift in migration at the earliest time points (15'; + indicates carboxypeptidase added). However, the presence of both mtRNAP and DNA (lanes 6, 12, 18, 24, and 30; green frames) protect the MTF1 C-terminus from digestion and no migration shift is seen in reactions with IC even after 240' digestion. Results are shown with the permission of Prof. Rui Sousa.

5.2. Interaction between the sc-mtRNAP and Mtf1

Previous attempts to identify the specific interaction surface between sc-mtRNAP and Mtf1 have failed. Deletions at the N-terminal and the C-terminal regions of Mtf1 resulted in the absence or decreased yield of IC (Cliften *et al.*, 1997) (Apendix 2), suggesting the involvement of multiple regions in sc-mtRNAP:Mtf1 interactions. However, the genetic screening has led to identification of point mutations that cause the lost or a drastic decrease in the mitochondrial transcription, and that are distributed in three clusters: region 1 (Y42C, H44P and L53H), region 2 (V135A, L154T and K157E) and region 3 (S218R, I221K, D225G) (Figure 40). Mapping of these regions in the X-ray structure of Mtf1 has revealed that they are facing different surfaces. Even more, several of them are located at the interior of the protein (region 1 and region 2), implying that they rather affect the structural integrity of the Mtf1 than the actual binding to sc-mtRNAP (Schubot *et al.*, 2001). The absence of any relevant structural data of the entire IC and previous mutational analysis has led to the proposal of false models that would explain their interaction (Paratkar y

Patel, 2010, Savkina *et al.*, 2010). Our EM study shows for the first time how these two proteins interact and, furthermore, excludes earlier hypothetic models, as the region 3 is placed at the opposite, outer face of Mtf1 (Figure 40, region 3 in green).

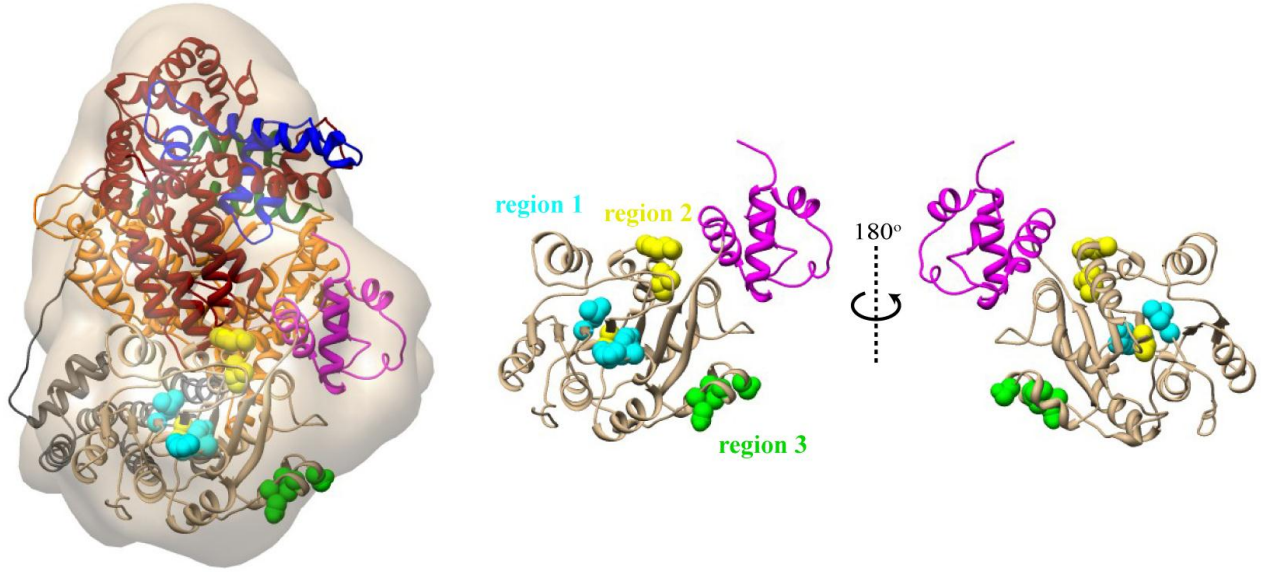


Figure 40: Interaction between sc-mtRNAP and Mtf1, and verification of previous mutational analysis of Mtf1. Docking of the X-ray atomic structures of human mtRNAP and Mtf1 in the cryo-3DEM density map of IC. The X-ray structures are color coded. Human mtRNAP: N-terminal extension (grey), NTD (orange), Thumb (dark green), Palm (red) and fingers (blue). Mtf1: N-terminal domain (tan) and C-terminal domain (magenta). The identified point mutations are grouped in three clusters: region 1 (Y42C, H44P and L53H) (cyan spheres), region 2 (V135A, L154T and K157E) (yellow spheres) and region 3 (S218R, I221K, D225G)) (green spheres).

Similar mutational analyses have identified three Rpo41 mutants (A631V, E1124K and K1273A) that restored the activity of Mtf1's V135A mutant (Cliften *et al.*, 1997). When marked on the fitted atomic structure of humant mtRNAP (Figure 41), only two of these mutations, A631V and E1124K, are placed at the proximity of Mtf1.

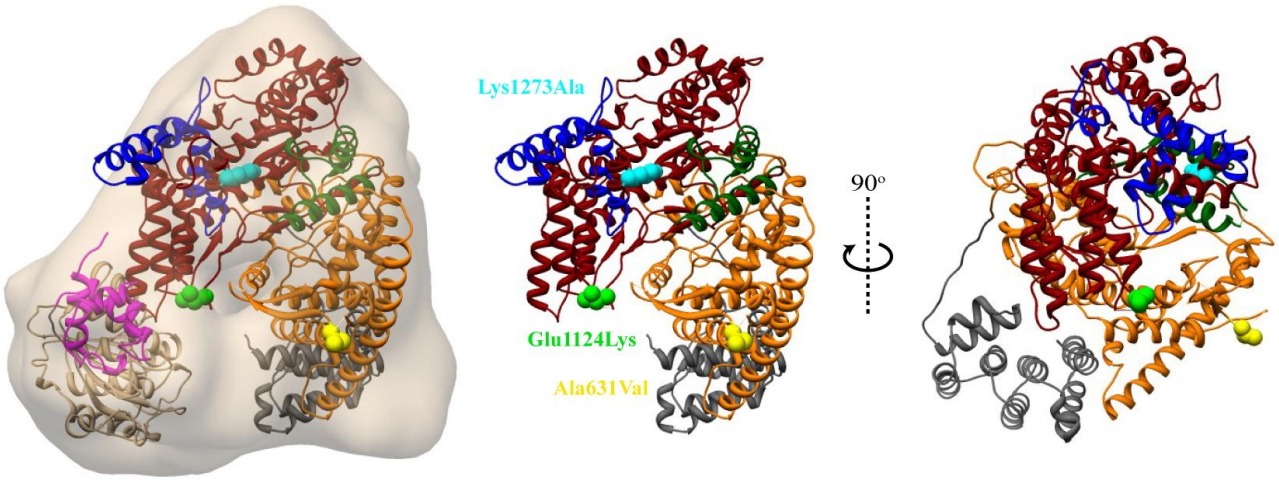


Figure 41: Interaction between sc-mtRNAP and Mtf1, and verification of previous mutational analysis of sc-mtRNAP. Docking of X-ray atomic structures of human mtRNAP and Mtf1 in the cryo-3DEM density map of IC. The X-ray structures are color coded. Human mtRNAP: N-terminal extension (grey), NTD (orange), Thumb (dark green), Palm (red) and fingers (blue). Mtf1: N-terminal domain (tan) and C-terminal domain (magenta). The homologues sequences of 7-10 residues long regions of sc-mtRNAP that included the point mutations were identified using the multireference alignment program Clustalw (http://www.genome.jp/tools/clustalw/). These corresponding homologue positions of A631V (yellow spheres) , E1124K (green spheres) and K1273A (cyan spheres) were marked on the corresponding atomic structure in order to see their position relative to the position of Mtf1 in the 3DEM model of IC.

It is important to notice that A631V and E1124K are adjacent to the intercalating β hairpin loop (residues 617-630) and the promoter specificity loop (residue 1127-1149), respectively (Deshpande y Patel, 2012). Thus, the Mtf1 might act on the repositioning of these elements, so they could participate in DNA melting. For instance, the position of the intercalating β hairpin in human homologue is not compatible with promoter melting as observed inT7 RNAP. It is suggested that this element assumes the right position upon the binding of h-mtTFB2 (Ringel *et al.*, 2011).

5.3. Transition from IC to EC

The transition from IC to EC in bacteria includes the successful tackling of two obstacles. First, during the initial steps of RNA synthesis, DNA gets scrunched, as the polymerase remains bound to the promoter and at the same time the downstream DNA is pulled into the active site. This results in the inevitable accumulation of free energy that eventually overcomes that of the promoter binding, which results in the breakage of holoenzyme-promoter contacts (Chen *et al.*, 2010). Second, once the growing RNA molecule reaches the length of 11-15 nts, it clashes into the σ3-4 linker and the σ4, which occupy the main and the RNA exit channel, respectively (Murakami *et al.*, 2002b). This triggers an array of chain events: the displacement of σ3.2 region from the holoenzyme, which destabilizes the interaction between the σ4 subdomain and β flap, causing the release of the UP element and the subsequent promoter escape (Kulbachinskiy y Mustaev, 2006).

The first notable conformational change in the initial stage of T7 RNAP's RNA synthesis occurs upon the addition of the 5th nts. The thumb subdomain bends towards the RNA and the template (Sousa y Mukherjee, 2003). As the RNA reaches the length of 7nts, the higher base paring energy makes the dissociation of DNA:RNA hybrid less probable. Upon the incorporation of the 8th base in the RNA, the PBD and the specificity loops undergo a left handed rotation of 45^{o} (Steitz, 2009), which is important for the accommodation of the growing DNA:RNA hybrid. This is followed by the additional bending of DNA, as it remains bound through the promoter region to the T7RNAP (Durniak *et al.*, 2008). The addition of the 9th nucleotide to RNA is followed by the collapse of the upstream edge of the transcription bubble (Martin *et al.*, 2005). This is the consequence of the combined effect of several factors: (1) the extension of the specificity loop over its limits and subsequent disruption of its contacts with the promoter, (2) competition of 3' end of RNA for the specificity loop and (3) the accumulation of the tension in the scrunched DNA (Sousa y Mukherjee, 2003). The transition to the stable and processive EC ends with the incorporation of the 12th nucleotide in the growing RNA molecule (Steitz, 2009, Tang *et al.*, 2009b). It includes rearrangements that mostly affect the N-terminal portion, causing the breakage of its interactions with the promoter. The N-terminal portion is also seen to be displaced for 35Å and rotated for 130^{o} relative to the core domain, which, however do not affect drastically the internal structure of this portion. The tip of the fingers domain moves towards the downstream DNA, contributing to the formation of the more extensive DNA binding site. The part of the thumb bends towards the fingers domain, which completes the formation of the pore that leads to

the active site (Tahirov *et al.*, 2002). The specificity loop and the H subdomain, which arises through the reorganization of NTD, form the RNA exit channel (Yin y Steitz, 2002).

To address the question how does the IC-EC transition occur in mtRNAP, our collaborators have generated the atomic model of IC, combining the information extracted from the corresponding atomic structures of T7RNAP and human mtRNAP and our 3DEM density maps of IC. They have used this atomic model to design the experiments with Fe-BABE nuclease, as described before. The nucleases of tethered ICs, which were halted at +0, +3, +7, or +22 (RNA length in nts), were activated by adding the ascorbic acid and hydrogen peroxide, and the digestion products were analyzed by electrophoresis. When the nuclease was tethered to R599C (N-terminal domain) of sc-mtRNAP, the initial cleavage was observed at the non-template strand with the cut centered at -11. This is exactly what the model predicts: R599 is centered on -11 of the non-template strand and the template strand is on the opposite side of the promoter and therefore protected from cutting by the nuclease at 599 (Figure 42A). When the RNA is extended to 7 nts, the cut site moves downstream to -7 and tracks the translocation of the polymerase. This indicates that sc-mtRNAP moves away from the promoter earlier than T7RNAP. When the nuclease is bound to the thumb (residue 780) or the fingers (residues 1025, 1074), a similar digestion pattern occurs in both IC and EC, which means that these two subdomains and the core domain in general maintain their position relative to the DNA (Figure 42A). This suggests that during the IC-EC transition, the sc-mtRNAP exhibits similar structural rearrangement as T7 RNAP, where the large conformational changes are constrained only to the N-terminal domain and the upstream region of DNA. Further support to this is given by comparisons between the 3DEM volumes of IC and EC. Taking into account that the catalytical core of the sc-mtRNAP should exhibit minimal rearrangements during the transition, the cryo-EM volumes of IC and EC have been superimposed by aligning the C-terminal portions of the fitted atomic structures of human homologue. Respective to that, the corresponding 3DEM volumes of the negatively stained complexes have been subsequently aligned. The resulting superimposition, which can be best observed when the 3DEM models of quasi-IC and EC are shown (Figure 42B), indicates the existence of an extra density that is occupied by the N-terminal extension and portion of the NTD of the atomic structure fitted into the 3DEM map of the EC. The quasi-IC can be used as it differs minimally from the electron density that corresponds to mtRNAP in IC (Figure 38A). The DNA digestion by the nuclease attached to thumb (residue 780) differs during the RNA extension from 3 to 6 nts (Figure 42A). This might reflect the movement of thumb subdomain and unwinding of the non-template strand, which has also been observed in T7 RNAP (Sousa y Mukherjee, 2003).

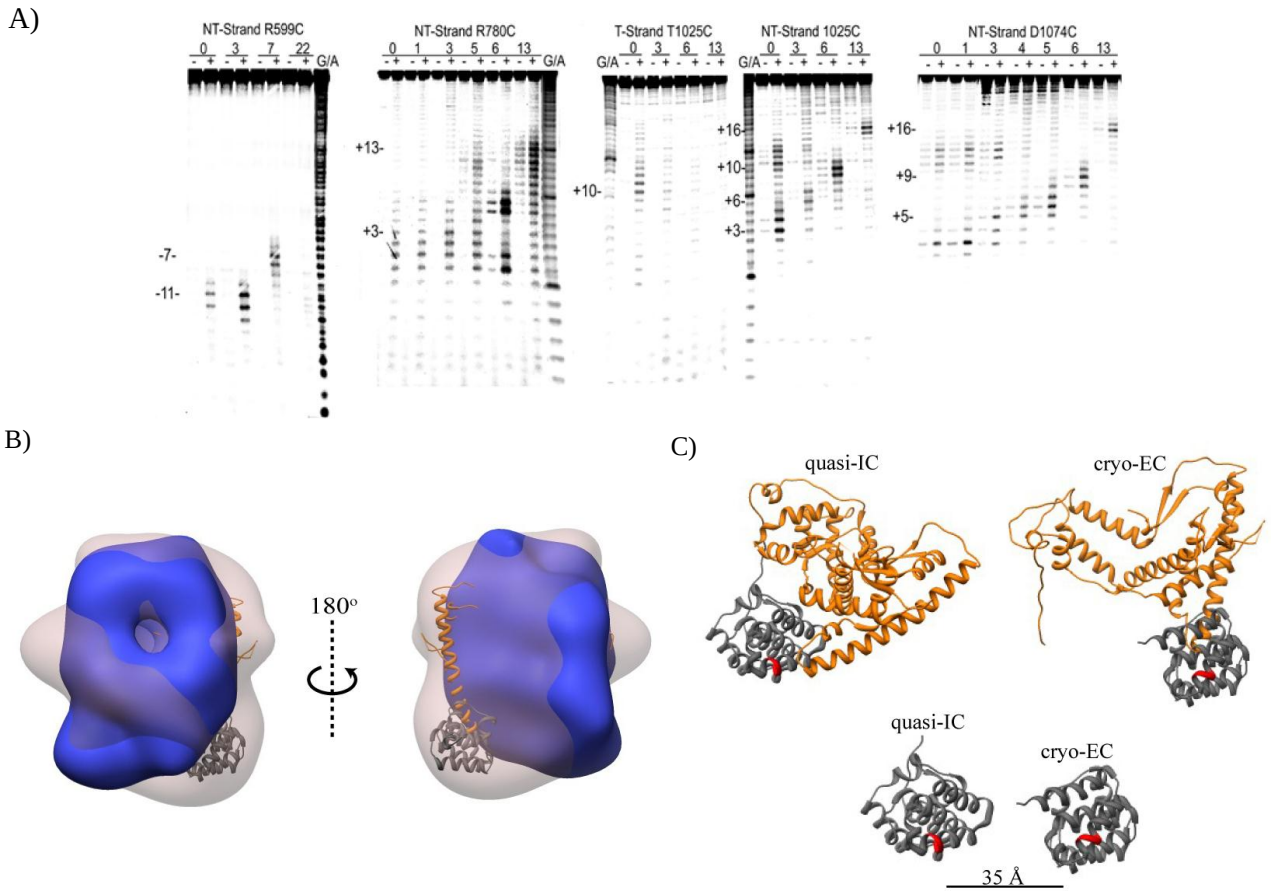


Figure 42: Transition from IC to EC- displacement of N-terminal portion of the sc-mtRNAP. A) The chemical nuclease cleavage pattern. When the nuclease is tethered to the residue at the position 599 (NTD) the DNA gets cut at the position -11, at the initial steps of RNA synthesis (RNA length 0-3). However, upon the incorporation of the 7th nucleotide this portion of the enzyme is displaced from the DNA and different cutting pattern can be observed, with the center of the cut at position -7. The DNA cleavage profiles of the nucleases attached to the thumb (residue 780) and finger domains (residues 1025, 1074) indicate that the C-terminal portion of the enzyme maintains the relative position towards the DNA in IC and EC. Results are shown with the permission of Prof. Rui Sousa. **B)** Superimposition of 3DEM volumes of negatively stained quasi-IC (blue) and EC (transparent) performed by aligning the core subdomain of the atomic structure (human homologue) that was fitted in the cryoEM volume of IC with the fragment of the core subdomain docked in the cryo-EM volume of EC. The X-ray structures shown are the N-terminal extension (grey) and the NTD (orange). There is minimal rotation of the catalytic core but the N-terminal domain exhibits the more notable displacement (as it was also seen in the T7RNAP). As there is no significant difference between the electron densities of sc-mtRNAP in 3DEM volumes of negatively stained quasi-IC and IC, except for the mass that is formed by upstream DNA and C-terminal tail of Mtf1 (Figures 24 and 28A), the comparison between the quasi-IC and EC is the most elegant way to show the transition observed by EM. **C)** The positions of the N-terminal portion in the two 3DEM volumes. The scale bar indicates the distance between the same residues in the two positions (red). Even though its value is in the range of the observed displacement in the T7 RNAP, it can be only used as an approximation. This is due to the low resolution of the EM study, so the exact positions of all the elements cannot be determined with a high precision

Further comparison between the volumes of streptavidin-tagged IC and EC shows the displacement of both termini of the DNA (Figure 43). The change in the position of the downstream end is within the range of being an artifact, while this cannot be said for the upstream one. This gives further support to the idea that IC-EC transition in T7 RNAP and sc-mtRNAP occurs in the same manner through a similar set of conformational changes.

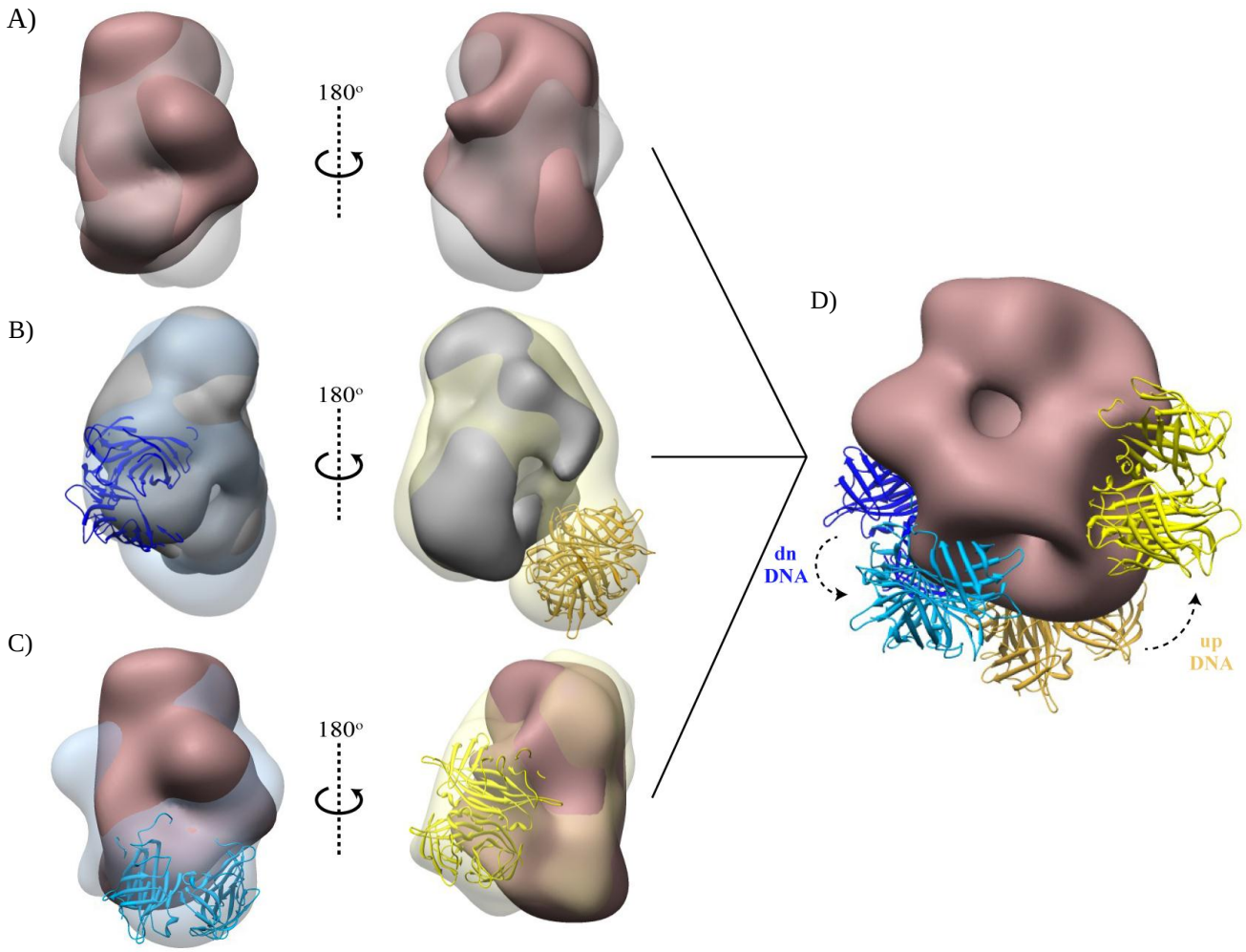


Figure 43: Transition from IC to EC- displacement of DNA. A) Superimposition of the IC (grey transparent) and EC (terracotta) shown in two lateral views in which the extra masses corresponding to streptavidins can be best shown. **B)** comparison between the 3DEM volumes of IC (grey) and downstream (transparent blue) and upstream labeled (transparent yellow) IC. The extra masses with the fitted atomic structures of streptavidin can be easily noted. The downstream and upstream positioned atomic structures of streptavidin are dyed in dark blue and dark yellow, respectively. **C)** This figure represents the same as the middle left except for the superimposition of EC (terracotta) and downstream and upstream labeled EC (terracotta, blue and yellow respectively). The corresponding atomic structures of streptavidin are colored in the lighter blue and yellow. **D)** Sum of all the observed positions of streptavidin, thus showing the change in the relative positions of upstream and downstream DNA termini during the IC-EC transition.

6. Conclusions

1. For the first time the three dimensional reconstructions of full length sc-mtRNAP in quasi-IC (bubble promoter:sc-mtRNAP complex), IC (bubble promoter:sc-mtRNAP complex:Mtf1)and EC (bubble DNA:RNA:sc-mtRNAP complex), together with the N-terminal deletion mutant EC, have been performed and the DNAs' termini have been localized in those complexes. Furthermore, the cryo-EM volumes of IC and EC have been obtained making them one of very few structures of similarly sized asymmetric complexes that have been characterized using cryo-EM.

2. The comparison between the EM volumes of streptavidin labeled quasi-IC and IC has confirmed the existence of the differential DNA bending that was previously shown to be responsible for discrimination of promoter from non-promoter sequences.

3. The transition from IC to EC starts earlier in the yeast mtRNAP than in the homologues T7 RNAP. Our structural data allowed the design of chemical nuclease experiments, which have shown that transition from IC to EC and the collapse of the upstream edge of the transcription bubble occurs upon the incorporation of the 7th nucleotide in the growing RNA molecule.

4. The comparison between the EM volumes of IC and EC and subsequently the streptavidin tagged complexes has confirmed that during the IC-EC transition, the sc-mtRNAP exhibits similar structural rearrangement as T7 RNAP, where the large conformational changes are constrained only to the N-terminal domain and the upstream region of DNA.

7. Appendix I

Image formation in electron microscope is based on the dual nature of electrons, and its consequence on their interactions with an examined sample. If we observe the electrons as the negatively charged elementary particles, upon their encounter with the specimen they can suffer different events that can be roughly classified as elastic and inelastic scattering events. Elastic scattering occurs upon the encounter of an incident electron with an atomic nucleus of a sample. Due to its negative charge, the incident electron gets attracted by the positively charged nucleus (Coulombic interaction). As the Coulumbic force ($F = Q1Q2 / 4\pi \varepsilon o\ r^2$) is directly proportional to the product of magnitude of each charge and inversely to the square of distance between them, the scattering angle will be bigger as the incident electron passes closer to the nucleus or as the encountered atom has a higher atomic number (Z) (S. Amelinckx, 1997). The momentum of the incident electrons is sufficient to avoid their capture by the nucleus. The elastically scattered electrons are characterized with no or negligible energy loss and the relatively large change in their trajectory, and a large scattering angle (1-100 miliradians). These electrons are removed by the objective apertures. Inelastic scattering results in energy loss (1-100 eV) and lower scattering angles (1-5 miliradians), thus making the objective apertures incapable of their removal, and contributing to background noise. These electrons transmit their energy to the sample, causing the formation of highly reactive ions and free radicals and subsequently irreversible damage of a sample (Amos *et al.*, 1982). The corpuscular nature of electron lies behind the amplitude component of image contrast, due to which the contrast is formed by the elimination of electrons with the high scattering angle (elastically scattered electrons). For unstained biological samples (cryo-EM) contribution of amplitude contrast to overall contrast is negligible.

The phase contrast, main component of total contrast, arises from the wave nature of electrons. Waves in the beams of any kind can be coherent or incoherent. The waves with the same wavelength and which are in the same phase are coherent waves, while the ones which differ in one or both of these characteristics are designated as incoherent waves. We can assume that the electron beam emitted from the electron source (without entering into the difference between the "conventional" cathode electron sources and field emission gun) represents a bundle of coherent waves- monochromatic beam. Upon the encounter with the samples diffraction occurs, generating both the coherent and incoherent waves, with approximate phase shift of $\pm \pi/2$. Waves interact with each other, what is known as interference. The interference of two waves with the same wavelength can result in:

- **Constructive interference:** If the waves are completely in phase with each other and have the same amplitude, then the amplitude of the resulting wave has the double the value of the original amplitudes.
- **Destructive interference**: If two waves with the same amplitude are exactly out of phase, meaning that the maximum of one wave is at the position of the minimum of the other, they are extinguished.

The phase contrast arises from the phase shift between the scattered and unscattered electron waves. This is the dominant contrast mechanism for thin stained and unstained specimens (Amos *et al.*, 1982). In

ideal situation, in the presence of aberrant-free electron microscope and the object perfectly being placed in focus, the phase contrast would be equal to zero (Fermat´s principle) (Amos *et al.*, 1982). Unfortunately, the reality is a drastically different. Namely, due to spherical (Cs) and chromatic aberrations (Cc) of the lenses, astigmatism (aberrance of perfectly symmetrical objective lens field), amplitude contrast, the image being out of focus (defocus levels), wave length of electrons and temporal and spatial coherence of the electron beam (Zhu *et al.*, 1997), there is a phase shift in the scattered wave. This phase shift is defined by the "contrast transfer function" or CTF (reference). The image recording process on film or CCD also contributes to the CTF (Koeck, 2000).

8. Appendix II

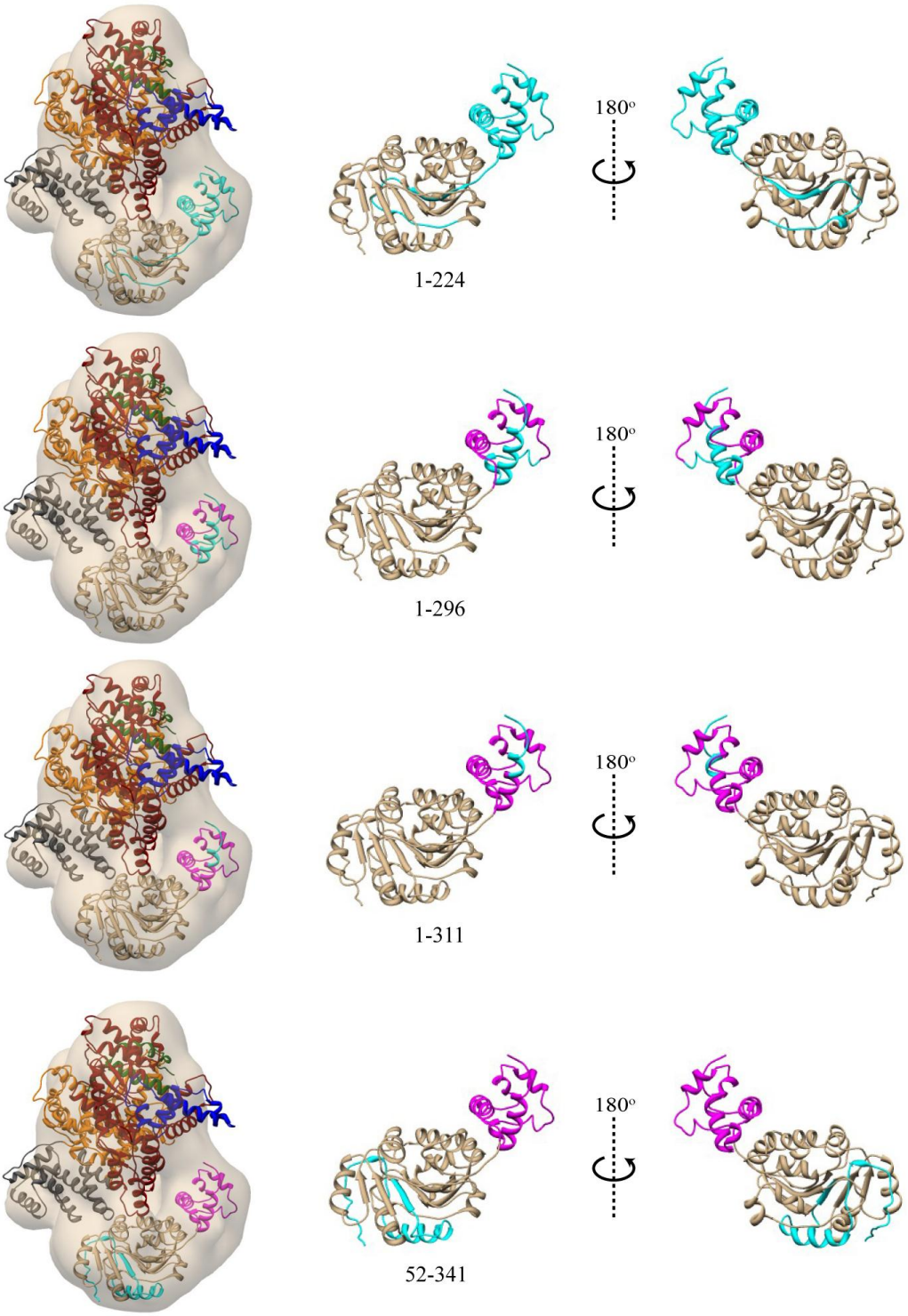


Figure 44: Interaction between sc-mtRNAP and Mtf1, and verification of deletion mutants of Mtf1. Docking of the X-ray atomic structures of human mtRNAP and Mtf1 in the cryo-3DEM density map of IC. The X-ray structures are color coded. Human mtRNAP: N-terminal extension (grey), NTD (orange), Thumb (dark green), Palm (red) and fingers (blue). Mtf1: N-terminal domain (tan) and C-terminal domain (magenta). The missing segments in deletion mutants are in cyan. The numbers bellow the each atomic model of MTF1 represents the remaining residues. Deletion of any of the segments resulted in the absence or drastically diminished capability of Mtf1 to interact with sc-mtRNAP.

9. RESUMEN

9.1. Introducción

En todos los organismos, la expresión génica requiere la presencia de máquinas moleculares complejas, las RNA polimerasas dependientes de DNA (RNAPs) que pueden dividirse en dos grupos: (1) RNAPs multiméricas, presentes en eubacterias, arqueas y en el núcleo de organismos eucariotas; y (2) RNAPs monoméricas, localizadas en algunos bacteriófagos, en mitocondrias y cloroplastos de organismos eucariotas, y en el núcleo de células de mamífero. Estos dos grupos presentan diferentes mecanismos o estrategias para abordar el proceso de síntesis de RNA, el cual requiere que la disociación de la polimerasa del molde de DNA sea mucho más lenta que la adición de bases a la cadena de RNA de nueva síntesis. Además, la RNAP debe tener la capacidad de translocarse a lo largo de la hebra molde. Para realizar esta función de manera eficiente, a lo largo de la evolución las RNAPs han adoptado soluciones basadas en la formación de estructuras que rodean el molde de DNA. En el caso de las RNAPs multiméricas, las diferentes subunidades constituyen una estructura de tipo "rosquilla", mientras que las RNAPs monoméricas son estructuras flexibles capaces de adaptar su conformación para envolver el molde (Bonner *et al.*, 1994, Sousa *et al.*, 1994). La RNAP del bacteriófago T7 (T7 RNAP), una proteína de 98 kDa, es el miembro mejor caracterizado de la familia de RNAPs monoméricas. Como el resto de RNAPs, ejerce tres funciones a lo largo de la transcripción: iniciación, elongación y terminación, y presenta dos estados activos: el modo de iniciación (o modo abortivo) y el modo de elongación (modo procesivo). Los estudios estructurales (Sousa *et al.*, 1993, Jeruzalmi y Steitz, 1998, Cheetham y Steitz, 1999, Yin y Steitz, 2002, Durniak *et al.*, 2008) revelan que la T7 RNAP muestra una arquitectura similar a la familia de DNA polimerasas I, ya que presenta una estructura secundaria y un plegamiento de sus dominios casi idéntico, a pesar de su escasa similitud de secuencia (Sousa y Mukherjee, 2003). La T7 RNAP presenta una estructura general en forma de mano derecha semicerrada, donde sus dominios, organizados alrededor de la ranura de unión al DNA, se nombran de acuerdo a los dominios del fragmento Klenow (el mayor producto proteolítico de la DNA polimerasa I):

- El **dominio del pulgar** (residuos 326-411), que está implicado en la estabilización del complejo de transcripción y en el mantenimiento de la procesividad durante la elongación (Sousa y Mukherjee, 2003)
- El **dominio de la palma** (residuos 412-553 y 785-883), que contiene dos horquillas β implicadas en las fases iniciales de la transcripción y cruciales para la actividad catalítica de la enzima (Jeruzalmi and Steitz 1998). Las horquillas β, formadas por los residuos 417-429, interaccionan con las bases 0-2 del híbrido de DNA-RNA, jugando un papel en su estabilización en las fases tempranas de la iniciación (Cheetham y Steitz, 1999, Imburgio *et al.*, 2002).
- El **dominio de los dedos** (residuos 554-784), responsable de la correcta selección de nucleótidos (NTPs) y de su discriminación con respecto a los desoxi-NTPs (dNTPs) (Jeruzalmi and Steitz, 1998). Este dominio también contiene una horquilla β que se extiende desde los residuos 739 a

770 y que se denomina lazo de especificidad debido a su papel en el reconocimiento y unión al promotor (Temiakov *et al.*, 2000).

- El **dominio N-terminal** (NTD, residuos 1-325), que contiene el lazo encargado de reconocer regiones ricas en AT (residuos 93-101) y el lazo de especificidad, que, junto con el lazo de especificidad del dominio de los dedos, son esenciales para el reconocimiento y unión al promotor, y por tanto para la iniciación de la transcripción (Muller *et al.*, 1988, Durniak *et al.*, 2008).

Las mitocondrias juegan un papel importante en una amplia variedad de procesos celulares, como la producción de energía, crecimiento y diferenciación celular, muerte celular programada (apoptosis), señalización e inmunidad innata (McWhirter *et al.*, 2005, Rehman, 2010, Mammucari *et al.*, 2011, Martinou y Youle, 2011, Cloonan y Choi, 2012). Por tanto, son una parte fundamental e indispensable de toda célula eucariota. La explicación más aceptada del origen de las mitocondrias la proporciona la teoría endosimbionte, según la cual las mitocondrias aparecieron hace 2 billones de años a partir de la invasión de α-proteobacterias ancestrales a arqueas o eucariotas primitivos (Andersson *et al.*, 1998). En base a esta teoría, podría esperarse que la expresión de los genes codificados en la mitocondria requiriera del empleo de una maquinaria transcripcional compleja. Sin embargo, la maquinaria de transcripción en mitocondrias es diferente de la maquinaria nuclear y no puede encuadrarse dentro de la familia de las RNAPs multiméricas. Se trata de un sistema mucho más sencillo que, en levaduras, consta de una única RNAP monomérica (producto del gen RPO41 o sc-mtRNAP) y de un factor de transcripción adicional denominado factor de transcripción mitocondrial 1 (Mtf1). La sc-mtRNAP, de 1351 residuos y un peso molecular de 153 kDa, contiene ocho regiones de homología con la T7 RNAP, y por tanto se incluye dentro de la familia de las RNAPs monoméricas, junto con todas las RNAPs mitocondriales. Estas regiones de homología constan de 9-89 residuos y presentan una identidad variable del 49-77%. En total, estas dos polimerasas comparten 219 residuos que suponen una identidad del 28% (Masters *et al.*, 1987). Esta evidente similitud de secuencia entre sc-mtRNAP y T7-RNAP permite inferir que esta homología se puede extender también a una homología estructural. La estructura atómica recientemente publicada de la mtRNAP humana confirma esta hipótesis (Ringel *et al.*, 2011). De esta manera, las mtRNAPs pueden dividirse en los siguientes dominios:

- **Dominio C-terminal**, donde se localizan la mayoría de las zonas de homología (residuos 648-1230) y que incluye los elementos responsables de establecer interacciones con la hebra molde del DNA, los NTPs y el Mg^{2+}, así como el sitio catalítico. Esto implica un mecanismo común de síntesis de RNA (incorporación de NTPs al RNA de nueva síntesis) entre las mtRNAPs y la T7 RNAP. El lazo de especificidad, implicado en el reconocimiento y apertura del promotor en la T7 RNAP, fue identificado en sc-mtRNAP (residuos 1127-1149) y en la mtRNAP humana (residuos 1086-1105) mediante análisis de secuencia.
- **Dominio N-terminal** (residuos 369-647), que posee una similitud de secuencia moderada con el dominio de unión a nucleótido (NBD) de la T7 RNAP, incluyendo el lazo de reconocimiento de regiones ricas en AT (residuos 481-497 en sc-mtRNAP y 450-470 en mtRNAP humana) y la

horquilla β intercalante (residuos 617-630 en sc-mtRNAP y 610-620 en mtRNAP humana) (Nayak *et al.*, 2009, Ringel *et al.*, 2011). Los estudios mutacionales realizados sobre cinco residuos localizados en el extremo del lazo de especificidad han mostrado la importancia de este elemento en la apertura del promotor, ya que impiden la transcripción a partir de DNA de cadena doble pero no del promotor parcialmente abierto. Sin embargo, se desconoce la función del lazo de reconocimiento de regiones ricas en AT en mtRNAPs, ya que los estudios mutacionales no han generado ningún cambio en la transcripción mitocondrial (Ringel *et al.*, 2011).

- **Extensión N-terminal**, que comprende los primeros 300 aminoácidos. Esta región es la más variable entre los distintos miembros de las mtRNAPs eucariotas. Existen diferentes hipótesis sobre la función de esta región. En primer lugar, los primeros 42 residuos de la mtRNAP humana constituyen la secuencia de señalización mitocondrial. En segundo lugar, diversos resultados a puntan a una función de los primeros 185 residuos en el paso de la transcripción a la traducción, ya que se ha descrito su interacción con proteínas implicadas en este proceso (Rodeheffer *et al.*, 2001, Bryan *et al.*, 2002, Rodeheffer y Shadel, 2003, Markov *et al.*, 2009). Además, se ha descrito que la eliminación de los primeros 270 residuos de la mtRNAP no impide la apertura del promotor y resulta en una menor acumulación de productos abortivos, lo que sugiere una función auto-inhibitoria de esta región (Paratkar *et al.*, 2011). Por último, la eliminación de los primeros 380 residuos impide a la sc-mtRNAP iniciar la transcripción a partir de DNA de cadena doble, pero no a partir del promotor parcialmente abierto, y también afecta la interacción con Mtf1, lo que indica que los residuos 270-380 están implicados en la interacción con Mtf1.

Mtf1 es una proteína compacta formada por: (1) un dominio α/β N-terminal, que contiene el sitio de unión a S-adenosil-L-metionina; (2) un dominio helicoidal C-terminal; y (3) una cola C-terminal implicada en la apertura del promotor y la iniciación de la transcripción, ya que interacciona con la hebra molde abierta (Deshpande y Patel, 2012). Varios estudios muestran que Mtf1 posee diferentes funciones en la iniciación, desde el reconocimiento específico del promotor, la introducción de una curvatura adicional en el DNA y la formación y estabilización del complejo de iniciación gracias a la interacción con los nucleótidos (nts) en las posiciones -10/-9 y -9/-8 de la hebra molde en la forma de doble cadena, y con los nts en las posiciones -5/-4 y +1/+2 de la hebra complementaria en la forma abierta, permitiendo así la selección del punto de inicio (Amiott y Jaehning, 2006, Paratkar y Patel, 2010).

9.2. Objetivos

El objetivo general de esta tesis es la caracterización estructural de la sc-mtRNAP, que permita obtener información sobre los mecanismos básicos de la transcripción mitocondrial en levaduras. En este contexto se plantearon los siguientes objetivos concretos:

1. Reconstrucción tridimensional de la sc-mtRNAP en el estado previo a la iniciación (quasi-IC) y localización del DNA en éste.

2. Reconstrucción tridimensional del complejo de iniciación (IC) e identificación de los extremos del DNA y de posibles sitios de interacción con Mtf1.
3. Reconstrucción tridimensional del complejo de elongación (EC) y determinación de la posición del DNA.
4. Localización del dominio N-terminal de la sc-mtRNAP mediante la reconstrucción tridimensional de mutantes de deleción, e identificación del DNA en estos complejos.

9.3. Materiales y Métodos

Los complejos proteicos analizados en esta tesis fueron proporcionados por el Prof. Rui Sousa (Departament of Biochemistry, The University of Texas Health Science Center at San Antonio, Texas, USA), mientras que los complejos marcados con estreptavidina se prepararon en nuestro laboratorio. Para ello, se mezclaron en relación equimolar oligonucleótidos marcados con biotina con las correspondientes hebras no marcadas parcialmente complementarias (hebras molde y complementaria de DNA en el caso de IC; hebras molde y complementaria de DNA junto con RNA en el caso de EC), incubados durante 3 min a 92ºC y enfriados lentamente a temperatura ambiente durante 1 h. Los complejos formados se mezclaron con estreptavidina en relación equimolar y se incubaron durante 45 min a temperatura ambiente. La pureza y estabilidad de los complejos se comprobó mediante análisis electroforéticos en geles desnaturalizantes y nativos de poliacrilamida (SDS-PAGE y Native-PAGE, respectivamente). Para la caracterización estructural de los complejos se utilizaron las técnicas de tinción negativa y criomicroscopía electrónica. Las imágenes de microscopía electrónica fueron tomadas en condiciones de mínima dosis. Las partículas fueron seleccionadas manualmente, extraídas de las micrografías y clasificadas utilizando los programas "maximum likelihood 2D" (ML2D) (Scheres *et al.,* 2005) y "clustering 2D" (CL2D) (Sorzano *et al.,* 2010). El uso combinado de estos dos métodos de clasificación se utilizó para obtener las referencias a partir de las cuales se generaron los modelos de referencia (modelos iniciales). Además, esta clasificación se usó para separar las partículas en las muestras más heterogéneas, como los complejos marcados con biotina-estreptavidina, y también para eliminar las partículas que puedan afectar etapas posteriores de la reconstrucción 3D, tales como las partículas que se encuentran muy próximas entre sí y/ o los agregados proteicos. Para solventar el problema de la heterogeneidad de estas muestras se empleó además el método de clasificación "KerDenSom" (Kernel Density Estimator Self-Organizing Map) (Pascual-Montano *et al.,* 2001). Este programa se basa en algoritmos de alineamiento de las partículas que permiten su separación en función de las diferencias presentes en una región concreta previamente definida por el usuario (Frank *et al.*, 1992, Penczek *et al.,* 1992). La mayoría de las etapas de reconstrucción 3D, incluyendo la generación de los modelos iniciales, se realizaron mediante el paquete de programas EMAN (Ludtke *et al.,* 1999). Existen diferentes metodologías implementadas en este paquete que pueden ser utilizadas para obtener mapas de referencia iniciales: (1) líneas comunes, basado en el teorema de la sección central (Crowther *et al.,* 1970), que postula que la transformada de Fourier de imágenes 2D, que representan las proyecciones del objeto observado, constituyen planos de ese objeto en el espacio de Fourier tridimensional; (2) blob, un objeto geométrico cuyas dimensiones deben corresponder a las del promedio de todas imágenes; y (3) ruido, basado

en el método anterior, ya que consiste en la asignación de ruido al blob generado previamente. Los modelos iniciales preliminares fueron refinados iterativamente frente a los datos experimentales en un proceso denominado refinamiento angular iterativo. El proceso comienza con la proyección del modelo inicial en 3D, donde el número de proyecciones depende del ángulo de proyección definido por el usuario. Las proyecciones son utilizadas como referencias iniciales (referencias teóricas) frente a las cuales se alinean y clasifican los datos experimentales. Después, las partículas clasificadas son promediadas, para crear así las referencias experimentales. Dado que las referencias experimentales contienen la información sobre su orientación espacial, a partir de las referencias teóricas de las que derivan, pueden ser sencillamente insertadas en el espacio de Fourier para crear un nuevo modelo, utilizado como modelo inicial para la siguiente iteración. Todo el proceso se repite hasta alcanzar un modelo estable, que es el que se produce cuando no se aprecian cambios significativos entre iteraciones (Ludtke et al., 1999). Una vez que las reconstrucciones independientes, obtenidas a partir de diferentes modelos iniciales, convergen a un mapa similar, las etapas finales del procesamiento se realizaron mediante "Projection Matching" (ajuste de las proyecciones) con el paquete XMIPP (Marabini et al., 1996). El modelo estable obtenido con EMAN se filtró a 60 Å y se usó como modelo inicial. En el caso de las muestras más heterogéneas (IC marcado con biotina-estreptavidina y crioME de EC) el modelo más estable se sometió a una clasificación con "maximum likelihood 3D" (ML3D) (Scheres, 2010). Las partículas separadas se sometieron de nuevo al ciclo completo de procesamiento con EMAN. La resolución de las reconstrucciones se determinó como el valor recíproco de FSC al cual el coeficiente de correlación es de 0.5. Antes de realizar los ajustes de la estructura atómica de la mtRNAP humana, los mapas tridimensionales se segmentaron utilizando el paquete de programas "Chimera" (Pettersen *et al.,* 2004). Los ajustes se realizaron manualmente y posteriormente se refinaron con Chimera utilizando un criterio de correlación en tiempo real entre el modelo de referencia (modelo final) y la superficie creada a partir de la estructura atómica.

9.4. Resultados y Discusión

La sc-mtRNAP exhibe una alta afinidad por el DNA, independientemente de su secuencia. Esto implica que la sc-mtRNAP no posee un mecanismo para distinguir la secuencia del promotor del DNA del resto de secuencias en la etapa inicial de unión al DNA. Por otro lado, la presencia de Mtf1 disminuye drásticamente la tasa de disociación de la sc-mtRNAP del DNA, sin afectar su tasa de asociación. Este efecto estabilizador de Mtf1 es más pronunciado cuando el complejo se forma con la secuencia promotora del DNA que con cualquier otra secuencia, aumentando cuatro veces la afinidad por el promotor (Tang *et al.,* 2009a). Sin embargo, esta diferencia es mínima cuando se compara con la T7 RNAP, que exhibe una afinidad 10^5 veces superior por el promotor que por otras secuencias (Bandwar y Patel, 2002). Esto implica la necesidad de un mecanismo diferente que permita al complejo sc-mtRNAP:Mtf1 discriminar entre las secuencias promotoras y no promotoras. Este nuevo mecanismo se basaría en los diferentes grados de curvatura del DNA, donde la presencia de Mtf1 induce un cambio en el ángulo de curvatura del promotor desde 52º a 89º, que no ocurre con secuencias no promotoras. Sólo los DNAs que presentan esta curvatura son posteriormente abiertos, lo que implica un acoplamiento entre ambos procesos (Deshpande and Patel, 2012).

Es importante destacar que un cambio similar en el ángulo de curvatura sucede también en el promotor de la T7 RNAP. En concreto, las interacciones entre el lazo de reconocimiento de regiones ricas en AT y el v de especificidad con la secuencia de DNA en las posiciones -17 a -13 y -11 a -7, respectivamente, provoca que el DNA se doble entre 40º y 60º. La incorporación del octavo nt en la molécula de RNA de nueva síntesis es seguida por una rotación a izquierdas del PBD y el lazo de especificidad de 45º (Steitz, 2009). Como el DNA permanece unido a la enzima, esta rotación induce un incremento adicional de la curvatura de 30º a 35º (Durniak *et al.*, 2008). Nuestros resultados estructurales están en total concordancia con estos datos. La mtRNAP y el DNA pueden interaccionar, como se ha descrito anteriormente, pero el complejo no es suficientemente estable para ser caracterizado por ME. En su lugar, puede formarse un complejo estable y transcripcionalmente activo con la burbuja del promotor, incluso en ausencia de Mtf1 (Matsunaga y Jaehning, 2004, Tang *et al.*, 2009a). La comparación de los volúmenes obtenidos para los complejos quasi-IC e IC revela la presencia de dos densidades electrónicas adicionales, la mayor de ellas situada en la parte frontal de la sc-mtRNAP y que corresponde a Mtf1, y otra más pequeña visible como una región que sobresale de la parte trasera de la sc-mtRNAP. Esta última región está compuesta parcialmente por el DNA, como se demostró mediante el marcaje del DNA corriente arriba del promotor. Además, la comparación entre los modelos tridimensionales de quasiIC e IC marcados corriente arriba muestra un desplazamiento significativo de este extremo del DNA entre los dos estados, de acuerdo con los diferentes grados de curvatura provocados en el DNA del promotor tras la unión con la sc-mtRNAP sola y en complejo con Mtf1 que se han descrito anteriormente.

El complejo sc-mtRNAP:Mtf1 abre el promotor desde el nt -4 al +2, por tanto la burbuja de transcripción inicial tiene un tamaño similar en levaduras y en el fago T7. En el caso de levaduras, la adición del primer nt en el RNA precede a un ligero reordenamiento estructural de la región que comprende los nts en las posiciones -2 y -1. Sin embargo, esto no provoca ningún cambio en el extremo corriente abajo de la burbuja de transcripción, que permanece en la posición +2 (Tang *et al.*, 2009a). Mtf1 en RPc y RPo parece establecer contactos con la región que se extiende desde la posición -10 a -8, lo cual sugiere su implicación en el reconocimiento de los elementos del promotor corriente arriba (Paratkar y Patel, 2010, Savkina *et al.*, 2010). Además, Mtf1 interacciona con los nts desde la posición -5 a -2 y desde +1 a +2 de la hebra complementaria, y con los nts +4 y +5 de la hebra molde (Paratkar *et al.*, 2011). Otros estudios (Savkina *et al.*, 2010) muestran la interacción de la cola C-terminal de Mtf1 (residuos 320-341) con los nts en las posiciones -4 y -3 de la hebra molde. Estos resultados parecen contradictorios, ya que la burbuja de transcripción empieza desde la posición -4. De esta manera, el C-terminal probablemente interviene en la formación del extremo ”upstream” de la burbuja de transcripción. Esto puede ocurrir a través de la estabilización de la cadena sencilla que queda expuesta como consecuencia de la curvatura del DNA y la acumulación de estrés. La interacción entre la cola C-terminal de Mtf1 y los nts en las posiciones -4 y -3 sugieren que esta región de Mtf1 queda oculta en la ranura de unión de la sc-mtRNAP. Nuestras reconstrucciones 3D de IC y los experimentos con una nucleasa química realizados por nuestros colaboradores en el laboratorio del Prof. Rui Sousa, diseñados en base a los resultados estructurales y

posterior ajuste, confirman esta hipótesis. Además, los experimentos con la carboxipeptidasa Y muestran que el C-terminal de Mtf1, bien aislada o en complejo con DNA, es rápidamente degradado a través de la eliminación de los 8 residuos C-terminales de este extremo (datos de espectrometría de masas, no mostrados). Sin embargo, cuando estos experimentos se realizan en presencia de IC, el C-terminal se encuentra muy protegido de la digestión. Esto indica que el C-terminal se oculta cuando Mtf1 se une al complejo IC invadiendo la ranura de unión de la polimerasa y la burbuja de DNA.

El resto de interacciones con la hebra complementaria del DNA hasta el nt en posición +2 pueden estar implicadas en la estabilización de RPo y en la apertura del promotor. Esto sugiere que Mtf1 tiene funciones similares a las del factor σ de las holoenzimas bacterianas. La proximidad de Mtf1 a la zona de transcripción y el fortalecimiento de la unión de la sc-mtRNAP y el primer nt incorporado en la molécula de RNA en su presencia indica que Mtf1 interviene en la unión de los dos primeros NTPs. Hasta la fecha no se ha logrado identificar con éxito la superficie de interacción entre sc-mtRNAP y Mtf1. La eliminación de las regiones N- y C-terminales de Mtf1 suponen la destrucción o disminución de la formación de IC (Cliften *et al.,* 1997), la implicación de múltiples regiones en la interacción con sc-mtRNAP. Sin embargo, el mapeo genético ha permitido la identificación de mutantes puntuales que provocan la pérdida o descenso drástico de la transcripción mitocondrial, todos ellos agrupados en tres regiones: la región 1(Y42C, H44P y L53H), la región 2 (V135A, L154T y K157E) y la región 3 (S218R, I221K y D225G). La localización de estas regiones en la estructura atómica de Mtf1 indica que están orientadas hacia diferentes superficies. Es más, algunas se localizan en el interior de la proteína (regiones 1 y 2), lo que implica que es posible que afecten a la integridad estructural de Mtf1 más que a su interacción con la sc-mtRNAP (Schubot *et al.*, 2001). La ausencia de datos estructurales relevantes del complejo IC completo y los análisis mutacionales previos han llevado a la hipótesis de modelos erróneos para intentar explicar esta interacción (Paratkar y Patel, 2010, Savkina *et al.*, 2010). Nuestro estudio por ME muestra por primera vez cómo estas dos proteínas interaccionan y además excluye modelos hipotéticos previos, ya que la región 3 está situada en la superficie opuesta, en la cara externa de Mtf1. El ajuste muestra que la ranura central de Mtf1, cargada positivamente y localizada entre el dominio α/β N-terminal y el dominio helicoidal C-terminal, se encuentra orientada hacia la sc-mtRNAP.

Para abordar la cuestión de cómo ocurre la transición entre IC y EC, nuestros colaboradores han generado un modelo atómico de IC, combinando la información extraída a partir de las estructuras atómicas de la T7 RNAP, de la mtRNAP humana y nuestros mapas de densidad de ME de IC. Este modelo se ha utilizado para diseñar experimentos de digestión con la nucleasa Fe-BABE. Los residuos indicados en el modelo atómico se sustituyeron por cisteínas, que posteriormente se conjugaron con la nucleasa mediante incubación de 1 h a 37ºC. El exceso de nucleasa se eliminó por filtración en gel, mientras que los mutantes unidos a la nucleasa se utilizaron para las reacciones de digestión. Las nucleasas unidas a los ICs, que fueron fijadas a +0, +3, +7 o +22 (longitud del RNA en nts), se activaron mediante la adición de ácido ascórbico y H_2O_2 y se detuvieron 5 s después utilizando un tampón de parada. Los productos de la reacción se analizaron mediante electroforesis. Cuando la nucleasa se encuentra unida al residuo R599C (dominio N-terminal) de

sc-mtRNAP, el centro de corte inicial se observó en la posición -11 de la hebra complementaria. Esto es exactamente lo que predice el modelo: R599 está centrado sobre la posición -11 de la hebra complementaria y la hebra molde se encuentra en el lado opuesto del promotor, por tanto protegida del corte de la nucleasa. Cuando el RNA se extiende a 7 nts, el sitio de corte se desplaza corriente abajo hasta la posición -7 y provoca la translocación de la polimerasa. Esto indica que la sc-mtRNAP se aleja del promotor más rápido que la T7 RNAP. Cuando la nucleasa está unida al pulgar (residuo 780) o a los dedos (residuos 1025 y 1074), se observa un patrón de digestión similar tanto en IC como en EC, lo que significa que estos dos subdominios y el dominio principal mantienen su posición relativa al DNA. Esto sugiere que durante la transición IC-EC la sc-mtRNAP muestra un reordenamiento estructural similar al de T7 RNAP, donde los grandes cambios estructurales se limitan al dominio N-terminal y la región corriente arriba del DNA. La comparación entre los volúmenes de los complejos IC y EC proporciona más evidencias en este sentido. Considerando que el núcleo catalítico de la sc-mtRNAP debe sufrir cambios estructurales mínimos en esta transición, los volúmenes de IC y EC obtenidos por crioME se superpusieron mediante el alineamiento de las porciones C-terminales de las estructuras atómicas ajustadas. Con este alineamiento como referencia, se superppusieron posteriormente los volúmenes obtenidos por tinción negativa. La superposición resultante indica la existencia de una densidad adicional ocupada por la extensión N-terminal y una porción del NTD de la estructura atómica ajustada en el mapa 3D de EC. La digestión del DNA por la nucleasa unida al pulgar (residuos 780) varía durante la extensión del RNA de 3 a 6 nts. Esto puede reflejar el movimiento del subdominio del pulgar y la apertura de la hebra complementaria, que también ha sido observada en T7 RNAP (Sousa y Mukherjee, 2003). La comparación adicional entre los volúmenes de IC y EC marcados con estreptavidina muestra el desplazamiento de ambos extremos del DNA. El cambio en la posición del DNA corriente abajo podría tratarse de un artefacto del proceso de reconstrucción 3D, mientras que este no puede ser el caso para el cambio observado en el extremo corriente arriba. Estos resultados apoyan la hipótesis de que la transición entre IC y EC en T7 RNAP y sc-mtRNAP ocurre de la misma forma y a través de un conjunto de cambios conformacionales similares.

9.5. Conclusiones

1. Por primera vez se ha obtenido la reconstrucción 3D de la sc-mtRNAP completa en sus estados quasiIC, IC y EC, además del mutante de eliminación N-terminal de EC, y se han localizado los extremos del DNA en estos complejos. Además, los volúmenes obtenidos por crioME de IC y EC constituyen unas de las escasas estructuras obtenidas de complejos asimétricos de tamaño similar utilizando esta técnica.
2. El marcaje con estreptavidina de quasi-IC e IC ha confirmado la existencia de una curvatura diferencial del DNA que había sido previamente descrita como la responsable de la discriminación de las secuencias promotoras y no promotoras.
3. La transición desde IC hasta EC comienza antes en la mtRNAP de levaduras que en su homóloga T7 RNAP. Nuestros datos estructurales han permitido el diseño de experimentos con una

nucleasa química, que han mostrado que la transición desde IC hasta EC y la destrucción del extremo ”upstream” de la burbuja de transcripción ocurren cuando se incorpora el séptimo nt a la molécula de RNA de nueva síntesis.

4. La sc-mtRNAP exhibe un reordenamiento estructural similar al de la T7 RNAP, en la que los cambios conformacionales más importantes tienen lugar en el dominio N-terminal y en la región ”upstream” del DNA.

REFERENCES:

Amiott, E. A. y Jaehning, J. A. (2006) Sensitivity of the yeast mitochondrial RNA polymerase to +1 and +2 initiating nucleotides. *J Biol Chem* **46**, 34982-8

Amos, L. A., Henderson, R. y Unwin, P. N. (1982) Three-dimensional structure determination by electron microscopy of two-dimensional crystals. *Prog Biophys Mol Biol* **3**, 183-231

Andersson, S. G., Zomorodipour, A., Andersson, J. O., Sicheritz-Ponten, T., Alsmark, U. C., Podowski, R. M., Naslund, A. K., Eriksson, A. S., Winkler, H. H. y Kurland, C. G. (1998) The genome sequence of Rickettsia prowazekii and the origin of mitochondria. *Nature* **6707**, 133-40

Aramayo, R., Sherman, M. B., Brownless, K., Lurz, R., Okorokov, A. L. y Orlova, E. V. (2011) Quaternary structure of the specific p53-DNA complex reveals the mechanism of p53 mutant dominance. *Nucleic Acids Res* **20**, 8960-71

Avery, O. T., Macleod, C. M. y McCarty, M. (1944) Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type Iii. *J Exp Med* **2**, 137-58

Balaji, S., Babu, M. M., Iyer, L. M., Luscombe, N. M. y Aravind, L. (2006) Comprehensive analysis of combinatorial regulation using the transcriptional regulatory network of yeast. *J Mol Biol* **1**, 213-27

Bandwar, R. P. y Patel, S. S. (2002) The energetics of consensus promoter opening by T7 RNA polymerase. *J Mol Biol* **1**, 63-72

Bar-Nahum, G. y Nudler, E. (2001) Isolation and characterization of sigma(70)-retaining transcription elongation complexes from Escherichia coli. *Cell* **4**, 443-51

Bartlett, M. S., Thomm, M. y Geiduschek, E. P. (2004) Topography of the euryarchaeal transcription initiation complex. *J Biol Chem* **7**, 5894-903

Bergstrom, C. T. y Pritchard, J. (1998) Germline bottlenecks and the evolutionary maintenance of mitochondrial genomes. *Genetics* **4**, 2135-46

Bonner, G., Lafer, E. M. y Sousa, R. (1994) The thumb subdomain of T7 RNA polymerase functions to stabilize the ternary complex during processive transcription. *J Biol Chem* **40**, 25129-36

Borukhov, S. y Nudler, E. (2003) RNA polymerase holoenzyme: structure, function and biological implications. *Curr Opin Microbiol* **2**, 93-100

Borukhov, S. y Nudler, E. (2008) RNA polymerase: the vehicle of transcription. *Trends Microbiol* **3**, 126-34

Bose, D., Pape, T., Burrows, P. C., Rappas, M., Wigneshweraraj, S. R., Buck, M. y Zhang, X. (2008) Organization of an activator-bound RNA polymerase holoenzyme. *Mol Cell* **3**, 337-46

Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 248-54

Bryan, A. C., Rodeheffer, M. S., Wearn, C. M. y Shadel, G. S. (2002) Sls1p is a membrane-bound regulator of transcription-coupled processes involved in Saccharomyces cerevisiae mitochondrial gene expression. *Genetics* **1**, 75-82

Camarero, J. A., Shekhtman, A., Campbell, E. A., Chlenov, M., Gruber, T. M., Bryant, D. A., Darst, S. A., Cowburn, D. y Muir, T. W. (2002) Autoregulation of a bacterial sigma factor explored by using segmental isotopic labeling and NMR. *Proc Natl Acad Sci U S A* **13**, 8536-41

Campbell, E. A., Muzzin, O., Chlenov, M., Sun, J. L., Olson, C. A., Weinman, O., Trester-Zedlitz, M. L. y Darst, S. A. (2002) Structure of the bacterial RNA polymerase promoter specificity sigma subunit. *Mol Cell* **3**, 527-39

Campbell, E. A., Pavlova, O., Zenkin, N., Leon, F., Irschik, H., Jansen, R., Severinov, K. y Darst, S. A. (2005) Structural, functional, and genetic analysis of sorangicin inhibition of bacterial RNA polymerase. *EMBO J* **4**, 674-82

Campbell, E. A., Westblade, L. F. y Darst, S. A. (2008) Regulation of bacterial RNA polymerase sigma factor activity: a structural perspective. *Curr Opin Microbiol* **2**, 121-7

Cashel, M., Hsu, L. M. y Hernandez, V. J. (2003) Changes in conserved region 3 of Escherichia coli sigma 70 reduce abortive transcription and enhance promoter escape. *J Biol Chem* **8**, 5539-47

Castrillo, J. I. y Oliver, S. G. (2004) Yeast as a touchstone in post-genomic research: strategies for integrative analysis in functional genomics. *J Biochem Mol Biol* **1**, 93-106

Cermakian, N., Ikeda, T. M., Cedergren, R. y Gray, M. W. (1996) Sequences homologous to yeast mitochondrial and bacteriophage T3 and T7 RNA polymerases are widespread throughout the eukaryotic lineage. *Nucleic Acids Res* **4**, 648-54

Cheetham, G. M., Jeruzalmi, D. y Steitz, T. A. (1999) Structural basis for initiation of transcription from an RNA polymerase-promoter complex. *Nature* **6731**, 80-3

Cheetham, G. M. y Steitz, T. A. (1999) Structure of a transcribing T7 RNA polymerase initiation complex. *Science* **5448**, 2305-9

Chen, J., Darst, S. A. y Thirumalai, D. (2010) Promoter melting triggered by bacterial RNA polymerase occurs in three steps. *Proc Natl Acad Sci U S A* **28**, 12523-8

Cheung, A. C., Sainsbury, S. y Cramer, P. (2011) Structural basis of initial RNA polymerase II transcription. *EMBO J* **23**, 4755-63

Ciampi, M. S. (2006) Rho-dependent terminators and transcription termination. *Microbiology* **Pt 9**, 2515-28

Cliften, P. F., Park, J. Y., Davis, B. P., Jang, S. H. y Jaehning, J. A. (1997) Identification of three regions essential for interaction between a sigma-like factor and core RNA polymerase. *Genes Dev* **21**, 2897-909

Cloonan, S. M. y Choi, A. M. (2012) Mitochondria: commanders of innate immunity and disease? *Curr Opin Immunol* **1**, 32-40

Cramer, P. (2002) Multisubunit RNA polymerases. *Curr Opin Struct Biol* **1**, 89-97

Cramer, P. (2004) Structure and function of RNA polymerase II. *Adv Protein Chem* 1-42

Cramer, P., Armache, K. J., Baumli, S., Benkert, S., Brueckner, F., Buchen, C., Damsma, G. E., Dengl, S., Geiger, S. R., Jasiak, A. J., Jawhari, A., Jennebach, S., Kamenski, T., Kettenberger, H., Kuhn, C. D., Lehmann, E., Leike, K., Sydow, J. F. y Vannini, A. (2008) Structure of eukaryotic RNA polymerases. *Annu Rev Biophys* 337-52

Crick, F. (1970) Central dogma of molecular biology. *Nature* **5258**, 561-3

Crowther, R. A., Amos, L. A., Finch, J. T., De Rosier, D. J. y Klug, A. (1970) Three dimensional reconstructions of spherical viruses by fourier synthesis from electron micrographs. *Nature* **5244**, 421-5

Deshpande, A. P. y Patel, S. S. (2012) Mechanism of transcription initiation by the yeast mitochondrial RNA polymerase. *Biochim Biophys Acta* **9-10**, 930-8

Dombroski, A. J., Walter, W. A. y Gross, C. A. (1993) Amino-terminal amino acids modulate sigma-factor DNA-binding activity. *Genes Dev* **12A**, 2446-55

Dombroski, A. J., Walter, W. A., Record, M. T., Jr., Siegele, D. A. y Gross, C. A. (1992) Polypeptides containing highly conserved regions of transcription initiation factor sigma 70 exhibit specificity of binding to promoter DNA. *Cell* **3**, 501-12

Durniak, K. J., Bailey, S. y Steitz, T. A. (2008) The structure of a transcribing T7 RNA polymerase in transition from initiation to elongation. *Science* **5901**, 553-7

Frank, J., Penczek, P. y Liu, W. (1992) Alignment, classification, and three-dimensional reconstruction of single particles embedded in ice. *Scanning Microsc Suppl* 11-20; discussion 20-2

Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H. y Oliver, S. G. (1996) Life with 6000 genes. *Science* **5287**, 546, 563-7

Gray, M. W. (1989) Origin and evolution of mitochondrial DNA. *Annu Rev Cell Biol* 25-50

Grimes, G. W., Mahler, H. R. y Perlman, R. S. (1974) Nuclear gene dosage effects on mitochondrial mass and DNA. *J Cell Biol* **3**, 565-74

Gusarov, I. y Nudler, E. (1999) The mechanism of intrinsic transcription termination. *Mol Cell* **4**, 495-504

Hershey, A. D. y Chase, M. (1952) Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J Gen Physiol* **1**, 39-56

Hirata, A., Klein, B. J. y Murakami, K. S. (2008) The X-ray crystal structure of RNA polymerase from Archaea. *Nature* **7180**, 851-4

Hirata, A. y Murakami, K. S. (2009) Archaeal RNA polymerase. *Curr Opin Struct Biol* **6**, 724-31

Holt, I. J., Harding, A. E. y Morgan-Hughes, J. A. (1988) Deletions of muscle mitochondrial DNA in patients with mitochondrial myopathies. *Nature* **6158**, 717-9

Huang, J., Brieba, L. G. y Sousa, R. (2000) Misincorporation by wild-type and mutant T7 RNA polymerases: identification of interactions that reduce misincorporation rates by stabilizing the catalytically incompetent open conformation. *Biochemistry* **38**, 11571-80

Huang, J. y Sousa, R. (2000) T7 RNA polymerase elongation complex structure and movement. *J Mol Biol* **3**, 347-58

Ikeda, R. A. (1992) The efficiency of promoter clearance distinguishes T7 class II and class III promoters. *J Biol Chem* **16**, 11322-8

Imburgio, D., Anikin, M. y McAllister, W. T. (2002) Effects of substitutions in a conserved DX(2)GR sequence motif, found in many DNA-dependent nucleotide polymerases, on transcription by T7 RNA polymerase. *J Mol Biol* **1**, 37-51

Ishihama, A., Fujita, N. y Glass, R. E. (1987) Subunit assembly and metabolic stability of E. coli RNA polymerase. *Proteins* **1**, 42-53

Jang, S. H. y Jaehning, J. A. (1991) The yeast mitochondrial RNA polymerase specificity factor, MTF1, is similar to bacterial sigma factors. *J Biol Chem* **33**, 22671-7

Jeon, Y. H., Negishi, T., Shirakawa, M., Yamazaki, T., Fujita, N., Ishihama, A. y Kyogoku, Y. (1995) Solution structure of the activator contact domain of the RNA polymerase alpha subunit. *Science* **5241**, 1495-7

Jeruzalmi, D. y Steitz, T. A. (1998) Structure of T7 RNA polymerase complexed to the transcriptional inhibitor T7 lysozyme. *EMBO J* **14**, 4101-13

Kim, H., Tang, G. Q., Patel, S. S. y Ha, T. (2012) Opening-closing dynamics of the mitochondrial transcription pre-initiation complex. *Nucleic Acids Res* **1**, 371-80

Koeck, P. J. (2000) Ins and outs of digital electron microscopy. *Microsc Res Tech* **3**, 217-23

Kravchenko, J. E., Rogozin, I. B., Koonin, E. V. y Chumakov, P. M. (2005) Transcription of mammalian messenger RNAs by a nuclear RNA polymerase of mitochondrial origin. *Nature* **7051**, 735-9

Kulbachinskiy, A. y Mustaev, A. (2006) Region 3.2 of the sigma subunit contributes to the binding of the 3'-initiating nucleotide in the RNA polymerase active center and facilitates promoter clearance during initiation. *J Biol Chem* **27**, 18273-6

Kurland, C. G. y Andersson, S. G. (2000) Origin and evolution of the mitochondrial proteome. *Microbiol Mol Biol Rev* **4**, 786-820

Kusser, A. G., Bertero, M. G., Naji, S., Becker, T., Thomm, M., Beckmann, R. y Cramer, P. (2008) Structure of an archaeal RNA polymerase. *J Mol Biol* **2**, 303-7

Li, W., Stevenson, C. E., Burton, N., Jakimowicz, P., Paget, M. S., Buttner, M. J., Lawson, D. M. y Kleanthous, C. (2002) Identification and structure of the anti-sigma factor-binding domain of the disulphide-stress regulated sigma factor sigma(R) from Streptomyces coelicolor. *J Mol Biol* **2**, 225-36

Li, Z., Wu, J. y Deleo, C. J. (2006) RNA damage and surveillance under oxidative stress. *IUBMB Life* **10**, 581-8

Lithgow, T. y Schneider, A. (2010) Evolution of macromolecular import pathways in mitochondria, hydrogenosomes and mitosomes. *Philos Trans R Soc Lond B Biol Sci* **1541**, 799-817

Lockshon, D., Zweifel, S. G., Freeman-Cook, L. L., Lorimer, H. E., Brewer, B. J. y Fangman, W. L. (1995) A role for recombination junctions in the segregation of mitochondrial DNA in yeast. *Cell* **6**, 947-55

Ludtke, S. J., Baldwin, P. R. y Chiu, W. (1999) EMAN: semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol* **1**, 82-97

Luft, R., Ikkos, D., Palmieri, G., Ernster, L. y Afzelius, B. (1962) A case of severe hypermetabolism of nonthyroid origin with a defect in the maintenance of mitochondrial respiratory control: a correlated clinical, biochemical, and morphological study. *J Clin Invest* 1776-804

MacAlpine, D. M., Perlman, P. S. y Butow, R. A. (2000) The numbers of individual mitochondrial DNA molecules and mitochondrial DNA nucleoids in yeast are co-regulated by the general amino acid control pathway. *EMBO J* **4**, 767-75

Malhotra, A., Severinova, E. y Darst, S. A. (1996) Crystal structure of a sigma 70 subunit fragment from E. coli RNA polymerase. *Cell* **1**, 127-36

Mammucari, C., Patron, M., Granatiero, V. y Rizzuto, R. (2011) Molecules and roles of mitochondrial calcium signaling. *Biofactors* **3**, 219-27

Mangus, D. A., Jang, S. H. y Jaehning, J. A. (1994) Release of the yeast mitochondrial RNA polymerase specificity factor from transcription complexes. *J Biol Chem* **42**, 26568-74

Marabini, R., Masegosa, I. M., San Martin, M. C., Marco, S., Fernandez, J. J., de la Fraga, L. G., Vaquerizo, C. y Carazo, J. M. (1996) Xmipp: An Image Processing Package for Electron Microscopy. *J Struct Biol* **1**, 237-40

Margulis, L. (1970) Recombination of non-chromosomal genes in Chlamydomonas: assortment of mitochondria and chloroplasts? *J Theor Biol* **2**, 337-42

Markov, D. A., Savkina, M., Anikin, M., Del Campo, M., Ecker, K., Lambowitz, A. M., De Gnore, J. P. y McAllister, W. T. (2009) Identification of proteins associated with the yeast mitochondrial RNA polymerase by tandem affinity purification. *Yeast* **8**, 423-40

Martin, C. T., Esposito, E. A., Theis, K. y Gong, P. (2005) Structure and function in promoter escape by T7 RNA polymerase. *Prog Nucleic Acid Res Mol Biol* 323-47

Martinez-Rucobo, F. W. y Cramer, P. (2012) Structural basis of transcription elongation. *Biochim Biophys Acta*

Martinou, J. C. y Youle, R. J. (2011) Mitochondria in apoptosis: Bcl-2 family members and mitochondrial dynamics. *Dev Cell* **1**, 92-101

Masters, B. S., Stohl, L. L. y Clayton, D. A. (1987) Yeast mitochondrial RNA polymerase is homologous to those encoded by bacteriophages T3 and T7. *Cell* **1**, 89-99

Matsunaga, M. y Jaehning, J. A. (2004) A mutation in the yeast mitochondrial core RNA polymerase, Rpo41, confers defects in both specificity factor interaction and promoter utilization. *J Biol Chem* **3**, 2012-9

Matsunaga, M., Jang, S. H. y Jaehning, J. A. (2004) Expression and purification of wild type and mutant forms of the yeast mitochondrial core RNA polymerase, Rpo41. *Protein Expr Purif* **1**, 126-30

McCulloch, V., Seidel-Rogol, B. L. y Shadel, G. S. (2002) A human mitochondrial transcription factor is related to RNA adenine methyltransferases and binds S-adenosylmethionine. *Mol Cell Biol* **4**, 1116-25

McQuillen, K., Roberts, R. B. y Britten, R. J. (1959) Synthesis of Nascent Protein by Ribosomes in Escherichia Coli. *Proc Natl Acad Sci U S A* **9**, 1437-47

McWhirter, S. M., Tenoever, B. R. y Maniatis, T. (2005) Connecting mitochondria and innate immunity. *Cell* **5**, 645-7

Mekler, V., Kortkhonjia, E., Mukhopadhyay, J., Knight, J., Revyakin, A., Kapanidis, A. N., Niu, W., Ebright, Y. W., Levy, R. y Ebright, R. H. (2002) Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex. *Cell* **5**, 599-614

Mendel, G. (1886) Versuche über Pflanzen-Hybriden. *Verh. Naturforsch* 3-47

Mindell, J. A. y Grigorieff, N. (2003) Accurate determination of local defocus and specimen tilt in electron microscopy. *J Struct Biol* **3**, 334-47

Muller, D. K., Martin, C. T. y Coleman, J. E. (1988) Processivity of proteolytically modified forms of T7 RNA polymerase. *Biochemistry* **15**, 5763-71

Murakami, E., Feng, J. Y., Lee, H., Hanes, J., Johnson, K. A. y Anderson, K. S. (2003) Characterization of novel reverse transcriptase and other RNA-associated catalytic activities by human DNA polymerase gamma: importance in mitochondrial DNA replication. *J Biol Chem* **38**, 36403-9

Murakami, K. S. y Darst, S. A. (2003) Bacterial RNA polymerases: the wholo story. *Curr Opin Struct Biol* **1**, 31-9

Murakami, K. S., Masuda, S., Campbell, E. A., Muzzin, O. y Darst, S. A. (2002a) Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science* **5571**, 1285-90

Murakami, K. S., Masuda, S. y Darst, S. A. (2002b) Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 A resolution. *Science* **5571**, 1280-4

Mustacchi, R., Hohmann, S. y Nielsen, J. (2006) Yeast systems biology to unravel the network of life. *Yeast* **3**, 227-38

Nayak, D., Guo, Q. y Sousa, R. (2009) A promoter recognition mechanism common to yeast mitochondrial and phage t7 RNA polymerases. *J Biol Chem* **20**, 13641-7

Nudler, E., Gusarov, I., Avetissova, E., Kozlov, M. y Goldfarb, A. (1998) Spatial organization of transcription elongation complex in Escherichia coli. *Science* **5375**, 424-8

Nudler, E., Mustaev, A., Lukhtanov, E. y Goldfarb, A. (1997) The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. *Cell* **1**, 33-41

Paget, M. S. y Helmann, J. D. (2003) The sigma70 family of sigma factors. *Genome Biol* **1**, 203

Paratkar, S., Deshpande, A. P., Tang, G. Q. y Patel, S. S. (2011) The N-terminal domain of the yeast mitochondrial RNA polymerase regulates multiple steps of transcription. *J Biol Chem* **18**, 16109-20

Paratkar, S. y Patel, S. S. (2010) Mitochondrial transcription factor Mtf1 traps the unwound non-template strand to facilitate open complex formation. *J Biol Chem* **6**, 3949-56

Pascual-Montano, A., Donate, L. E., Valle, M., Barcena, M., Pascual-Marqui, R. D. y Carazo, J. M. (2001) A novel neural network technique for analysis and classification of EM single-particle images. *J Struct Biol* **2-3**, 233-45

Penczek, P., Radermacher, M. y Frank, J. (1992) Three-dimensional reconstruction of single particles embedded in ice. *Ultramicroscopy* **1**, 33-53

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. y Ferrin, T. E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **13**, 1605-12

Rehman, J. (2010) Empowering self-renewal and differentiation: the role of mitochondria in stem cells. *J Mol Med (Berl)* **10**, 981-6

Ringel, R., Sologub, M., Morozov, Y. I., Litonin, D., Cramer, P. y Temiakov, D. (2011) Structure of human mitochondrial RNA polymerase. *Nature* **7368**, 269-73

Rodeheffer, M. S., Boone, B. E., Bryan, A. C. y Shadel, G. S. (2001) Nam1p, a protein involved in RNA processing and translation, is coupled to transcription through an interaction with yeast mitochondrial RNA polymerase. *J Biol Chem* **11**, 8616-22

Rodeheffer, M. S. y Shadel, G. S. (2003) Multiple interactions involving the amino-terminal domain of yeast mtRNA polymerase determine the efficiency of mitochondrial protein synthesis. *J Biol Chem* **20**, 18695-701

Rogowska, A. T., Puchta, O., Czarnecka, A. M., Kaniak, A., Stepien, P. P. y Golik, P. (2006) Balance between transcription and RNA degradation is vital for Saccharomyces cerevisiae mitochondria: reduced transcription rescues the phenotype of deficient RNA degradation. *Mol Biol Cell* **3**, 1184-93

Rong, M., He, B., McAllister, W. T. y Durbin, R. K. (1998) Promoter specificity determinants of T7 RNA polymerase. *Proc Natl Acad Sci U S A* **2**, 515-9

Savkina, M., Temiakov, D., McAllister, W. T. y Anikin, M. (2010) Multiple functions of yeast mitochondrial transcription factor Mtf1p during initiation. *J Biol Chem* **6**, 3957-64

Scheres, S. H. (2010) Classification of structural heterogeneity by maximum-likelihood methods. *Methods Enzymol* 295-320

Scheres, S. H., Valle, M., Nunez, R., Sorzano, C. O., Marabini, R., Herman, G. T. y Carazo, J. M. (2005) Maximum-likelihood multi-reference refinement for electron microscopy images. *J Mol Biol* **1**, 139-49

Schubot, F. D., Chen, C. J., Rose, J. P., Dailey, T. A., Dailey, H. A. y Wang, B. C. (2001) Crystal structure of the transcription factor sc-mtTFB offers insights into mitochondrial transcription. *Protein Sci* **10**, 1980-8

Schwartz, E. C., Shekhtman, A., Dutta, K., Pratt, M. R., Cowburn, D., Darst, S. y Muir, T. W. (2008) A full-length group 1 bacterial sigma factor adopts a compact structure incompatible with DNA binding. *Chem Biol* **10**, 1091-103

Shadel, G. S. y Clayton, D. A. (1995) A Saccharomyces cerevisiae mitochondrial transcription factor, sc-mtTFB, shares features with sigma factors but is functionally distinct. *Mol Cell Biol* **4**, 2101-8

Shutt, T. E. y Gray, M. W. (2006) Homologs of mitochondrial transcription factor B, sparsely distributed within the eukaryotic radiation, are likely derived from the dimethyladenosine methyltransferase of the mitochondrial endosymbiont. *Mol Biol Evol* **6**, 1169-79

Sorzano, C. O., Bilbao-Castro, J. R., Shkolnisky, Y., Alcorlo, M., Melero, R., Caffarena-Fernandez, G., Li, M., Xu, G., Marabini, R. y Carazo, J. M. (2010) A clustering approach to multireference alignment of single-particle projections in electron microscopy. *J Struct Biol* **2**, 197-206

Sousa, R., Chung, Y. J., Rose, J. P. y Wang, B. C. (1993) Crystal structure of bacteriophage T7 RNA polymerase at 3.3 A resolution. *Nature* **6438**, 593-9

Sousa, R. y Mukherjee, S. (2003) T7 RNA polymerase. *Prog Nucleic Acid Res Mol Biol* 1-41

Sousa, R., Rose, J. y Wang, B. C. (1994) The thumb's knuckle. Flexibility in the thumb subdomain of T7 RNA polymerase is revealed by the structure of a chimeric T7/T3 RNA polymerase. *J Mol Biol* **1**, 6-12

Steitz, T. A. (2009) The structural changes of T7 RNA polymerase from transcription initiation to elongation. *Curr Opin Struct Biol* **6**, 683-90

Tahirov, T. H., Temiakov, D., Anikin, M., Patlan, V., McAllister, W. T., Vassylyev, D. G. y Yokoyama, S. (2002) Structure of a T7 RNA polymerase elongation complex at 2.9 A resolution. *Nature* **6911**, 43-50

Tang, G. Q., Paratkar, S. y Patel, S. S. (2009a) Fluorescence mapping of the open complex of yeast mitochondrial RNA polymerase. *J Biol Chem* **9**, 5514-22

Tang, G. Q., Roy, R., Bandwar, R. P., Ha, T. y Patel, S. S. (2009b) Real-time observation of the transition from transcription initiation to elongation of the RNA polymerase. *Proc Natl Acad Sci U S A* **52**, 22175-80

Temiakov, D., Mentesana, P. E., Ma, K., Mustaev, A., Borukhov, S. y McAllister, W. T. (2000) The specificity loop of T7 RNA polymerase interacts first with the promoter and then with the elongating transcript, suggesting a mechanism for promoter clearance. *Proc Natl Acad Sci U S A* **26**, 14109-14

Thorsness, P. E. y Fox, T. D. (1990) Escape of DNA from mitochondria to the nucleus in Saccharomyces cerevisiae. *Nature* **6282**, 376-9

Thorsness, P. E. y Fox, T. D. (1993) Nuclear mutations in Saccharomyces cerevisiae that affect the escape of DNA from mitochondria to the nucleus. *Genetics* **1**, 21-8

Thuman-Commike, P. A. (2001) Single particle macromolecular structure determination via electron microscopy. *FEBS Lett* **2**, 199-205

Timmis, J. N., Ayliffe, M. A., Huang, C. Y. y Martin, W. (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat Rev Genet* **2**, 123-35

Treutlein, B., Muschielok, A., Andrecka, J., Jawhari, A., Buchen, C., Kostrewa, D., Hog, F., Cramer, P. y Michaelis, J. (2012) Dynamic architecture of a minimal RNA polymerase II open promoter complex. *Mol Cell* **2**, 136-46

van Heel, M. (1984) Multivariate statistical classification of noisy images (randomly oriented biological macromolecules). *Ultramicroscopy* **1-2**, 165-83

Vassylyev, D. G., Sekine, S., Laptenko, O., Lee, J., Vassylyeva, M. N., Borukhov, S. y Yokoyama, S. (2002) Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution. *Nature* **6890**, 712-9

Vassylyev, D. G., Vassylyeva, M. N., Perederina, A., Tahirov, T. H. y Artsimovitch, I. (2007a) Structural basis for transcription elongation by bacterial RNA polymerase. *Nature* **7150**, 157-62

Vassylyev, D. G., Vassylyeva, M. N., Zhang, J., Palangat, M., Artsimovitch, I. y Landick, R. (2007b) Structural basis for substrate loading in bacterial RNA polymerase. *Nature* **7150**, 163-8

von Hippel, P. H. (1998) An integrated model of the transcription complex in elongation, termination, and editing. *Science* **5377**, 660-5

Wallace, D. C. (1992) Diseases of the mitochondrial DNA. *Annu Rev Biochem* 1175-212

Wallace, D. C., Lott, M. T., Shoffner, J. M. y Brown, M. D. (1992) Diseases resulting from mitochondrial DNA point mutations. *J Inherit Metab Dis* **4**, 472-9

Werner, F., Eloranta, J. J. y Weinzierl, R. O. (2000) Archaeal RNA polymerase subunits F and P are bona fide homologs of eukaryotic RPB4 and RPB12. *Nucleic Acids Res* **21**, 4299-305

Wu, C. C., Herzog, F., Jennebach, S., Lin, Y. C., Pai, C. Y., Aebersold, R., Cramer, P. y Chen, H. T. (2012) RNA polymerase III subunit architecture and implications for open promoter complex formation. *Proc Natl Acad Sci U S A*

Yin, Y. W. y Steitz, T. A. (2002) Structural basis for the transition from initiation to elongation transcription in T7 RNA polymerase. *Science* **5597**, 1387-95

Zeviani, M., Bresolin, N., Gellera, C., Bordoni, A., Pannacci, M., Amati, P., Moggio, M., Servidei, S., Scarlato, G. y DiDonato, S. (1990) Nucleus-driven multiple large-scale deletions of the human mitochondrial genome: a new autosomal dominant disease. *Am J Hum Genet* **6**, 904-14

Zhang, G., Campbell, E. A., Minakhin, L., Richter, C., Severinov, K. y Darst, S. A. (1999) Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 A resolution. *Cell* **6**, 811-24

Zhu, A. Q. y von Hippel, P. H. (1998) Rho-dependent termination within the trp t' terminator. I. Effects of rho loading and template sequence. *Biochemistry* **32**, 11202-14

Zhu, J., Penczek, P. A., Schroder, R. y Frank, J. (1997) Three-dimensional reconstruction with contrast transfer function correction from energy-filtered cryoelectron micrographs: procedure and application to the 70S Escherichia coli ribosome. *J Struct Biol* **3**, 197-219